

行政院國家科學委員會專題研究計畫 成果報告

問題為基礎的學習對學生學習方式改變的評估研究(2/2)

計畫類別：個別型計畫

計畫編號：NSC93-2516-S-002-002-

執行期間：93年08月01日至94年07月31日

執行單位：國立臺灣大學醫學院家庭醫學科

計畫主持人：梁繼權

共同主持人：呂碧鴻，王維典

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 94 年 12 月 15 日

中文摘要

國內有關 PBL 的學生學習表現的評估研究並不多，在國外研究中對於 PBL 的學習效果是否可以加強學生的理解能力，亦缺乏有力的研究結果支持。因此在進一步推廣 PBL 教學的時候，需要更多的研究來作為課程設計與執行時的參考。本研究目的包括：1.針對台大醫學院 PBL 課程的實施狀況設計合適的學習評估方法；2.根據 PBL 的學習理論評估目標學習行為的達成情況；3.建立相關的評估工具與流程；4.探討學習評估對學生整體成績的預測效度(predictive validity)及學生自評與教師評估的相關關係；5.以概念圖作為學生理解能力評估工具的可行性及相關因素；6.研究 PBL 是否影響學生理解能力。

本研究以台大醫學系二至四年級學生及小班教學老師為 PBL 學生能力表現評估工具—Tutotest^C的研究對象。並作中文版 Tutotest^C的信效度研究。第二部分以六、七年級實習醫師為概念圖繪製及評估的研究對象。發展概念圖的施測流程及評議分方法，同時評估評估者一致性及概念圖與相關因素的關係。

研究結果顯示中文版 Tutotest^C具有良好的內部一致性信度與兩週的再測信度，因素分析結果與原英文問卷相似，具有良好的建構效度。學生自評與老師評估相關度低，顯示學生可能缺乏良好的自我評估能力，與國外研究結果十分相似。然而老師的評分偏差亦需同時考慮。老師採用整體評量與用 Tutotest^C評估雖有相關但仍不夠高，顯示不同評分方法會影響老師的評分結果，老師對學生的整體印象與實際的行為表現有差距。

概念圖的研究結果顯示具有執行的可能性，但不同評分方式的評估者間一致性有極大差異。以 holistic 法及 relational 法具有較佳之評估者間一致性。學生繪製的概念圖的複雜度亦有很大差異。概念圖的評分與臨床實習年資、是否有 PBL 經驗及實習成績無關，但因研究樣本太少，樣本選取可能有偏差，仍需進一步澄清。

關鍵詞：問題為基礎的小班教學、教學評估、概念圖

Abstract

This study validated the Tutotest in a hybrid problem-based learning (PBL) curriculum and evaluated the agreement of student's self-assessment and tutors' assessment. A Chinese version of the Tutotest (Tutotest-C) was developed and applied to the summative evaluation of the second to fourth year medical students taking hybrid problem-based learning (PBL) courses in a seven-year curriculum medical school.

Forty-four tutors and students completed 370 evaluations at the end of the first semester in 2004. There was significant correlation between global rating and Tutotest-C ($r=0.44$, $p<0.001$). The Cronbach's α coefficient was 0.97. Two-week test-retest correlation coefficient was 0.85. Factor analysis revealed five factors, where four of these factors were similar to the factors of "effectiveness in group", "communication and leadership skills", and "respect for others" identified in the original Tutotest. "Hypothesis forming and testing" instead of "scientific curiosity" became the fifth factor in our data.

Our study validated the Tutotest-C in a hybrid PBL curriculum and students from Chinese education system. Test-retest reliability measure with a two-week interval at the end of the PBL tutorial confirmed the stability of the Tutotest, which has not been reported. Since most Asia medical schools adopted a hybrid PBL curriculum, validation of the Tutotest in a hybrid PBL curriculum has practical implications. On the other hands, students' self-assessment were not related to tutors' rating. Low correlations between students and tutors in three Tutorest subscales: the "effectiveness in group", "ability to master PBL", and "communication and leadership skills" were observed in the third year students. Moderate correlation between tutors' global ratings and Tutotest rating was noted that implied different aspects of ratings.

In the concept map study, sixth and seventh year medical students who attained the family medicine clinical elective were included in this study. In this study, a very common concept "headache" was chosen and students were asked to complete their map in one hour with referring to their books or other information. A brief focus interview was conduct immediately after students completed their maps. Six family medicine faculties were trained to score the concept maps using four scoring methods. Each concept map was randomly assigned to score by two faculties independently. Each faculty score each map twice using two different scoring methods. As a result, each concept map was scored by two different faculties with four scoring methods.

Thirty-three concept maps on a topic of headache were collected at the end of the study period. All students could complete their map within the time limit of one hour. There were very wide ranges and large standard deviations of score among different scoring methods especially in scores using Novak & Gowin, and Markham method. Inter-rater correlations were very low except moderate correlations between the holistic and relational methods. Our results indicated that concept mapping could be applied in

medical education to evaluate conceptual understandings of medical knowledge. However, there were significant differences in inter-rater reliability among different scoring methods. Holistic and relational methods of scoring have acceptable inter-rater reliabilities in the evaluation of the concept maps. Limitation of our study included the use of a group of very special subjects and of a very small sample size. Further research should be done on a large and heterogeneous sample in order to evaluate the effect of using concept map in medical education.

Key Word: Problem-based learning, evaluation, concept mapping

前言

當 PBL 的熱潮逐漸在世界各地擴散之際，有關 PBL 對學習的成效便更加受到重視，由於 PBL 課程比傳統課程需要耗費更多的資源，因此大家便更關心 PBL 所能帶來的成本效益。

Colliver 整理 1992 至 1998 在 8 本醫學教育研究最常發表的雜誌，找出比較 PBL 與傳統教學差異的文章共八篇進行 meta-analysis，作者發現這些研究都無法提出有力的證明，顯示 PBL 比傳統教學對學生有較佳的學習成果。為何會導致上述結果，Colliver 認為有下列的可能原因：第一、學習的情境理論認為在 A 的情境下學習，在 A 的情境下較能回憶與應用，Colliver 懷疑 PBL 與傳統學習的學習情境其實沒有太大的差異，而且兩者與將來行醫時應用的情境都不同。第二、過去知識的活化可以幫助學習新的事物，使之與過去知識連結以方便以後知識的提取。Colliver 認為如果傳統課程學生與 PBL 學生花同樣多的時間去學習的話，兩者在知識活化的效果上便可能會沒有差異。此外，PBL 的另一個目的是養成終身自我學習能力，但只有一篇研究發現 PBL 學校的畢業生在行醫時較其他傳統教學學校畢業生能應用高血壓的最新治療標準，但由於 PBL 教學學校多為在心血管研究領域知名學校，因此 Colliver 對於結果存保留的態度，最後，Colliver 認為 PBL 的效果研究仍然不足。

Colliver 的論文在學術界引起一番爭論，Albanese 稍後發表文章提出不同的觀點，他認為 Colliver 對於“effect size”的要求太高，教育效果的改變有限，不可以用太高的“effect size”來衡量，醫學生是一群十分獨特的群體，在經歷重重考驗才能進入醫學院，他們是傳統教學法所製造出來的頂尖產品，因此 PBL 教學可能產生多大的改變，Albanese 是保持保守的態度。另一方面，目前醫學院之體制是為了傳統式教學而設計的，一個 PBL 教學課程建立在傳統的體制上使得效果不易在短期內彰顯。在教學效果的評估上，Albanese 認為用國家考試的結果來評估 PBL 的效果是不當的，因為它們是為了傳統式教學的評估而設計的。Albanese 指出 PBL 並非如 Colliver 所言缺乏有力的理論基礎，除了情境理論外(Albanese 同意這是 PBL 比較薄弱的地方)，其他如訊息處理理論(information-processing theory)、合作學習(cooperative learning)、自我決定理論(self-determinative theory)、控制理論(control theory)等都是 PBL 的理論基礎。作者最後指出就算 PBL 對於增加醫學知識與技巧沒有助益，最少可以為學生與老師營造一個較佳、較人性化的學習環境。

McMaster 大學的 Norman 與 Schmidt 對 Colliver 對 PBL 的評論做出回應，他們認為學習情境是很難去用實驗方法去控制的，大部分的研究都受到干擾變項的影響，他們亦認為 Colliver 並沒有把重要的與知識活化的研究放進他的研究之中。Norman 與 Schmidt 進一步指出利用藥物研究的大型臨床試驗方法來做教育研究是不當的，因為教育研究無法達到「盲目介入」(blinded intervention)，而且無法獲得單純的結果，而且不能確保每一個受試者都獲得相同「劑量」的介入，作者認為目前必須進行更多的學習理論層面及計畫層面的研究評估，而非用目前的研究結果去否定 PBL 的價值。

國內有關 PBL 的研究並不多，在研究的深度上仍在起步階段，因此在進一步推廣 PBL 教學的時候，需要更多的研究來作為課程設計與執行時的參考。過去的教學評估，尤其是 PBL 課程的評估部分，為課程的滿意度評估(呂碧鴻等 1998，陳震寰等 1998，賴玉玲等 2000，陳思光等 2000)、可行性評估(陳震寰等 1998，賴玉玲

2000, 陳玉琨等 2001)、課程的整體評估(梁繼權等 1997, 梁繼權等 1999)、評估工具的研究(高美英等 1997)、教師培訓計劃的評估(梁繼權等 1997)、教學方式評估(梁繼權等 1998)、學生在小組學習的表現評估(謝明憲等 1999, 潘愛琬等 2000)。

縱觀國內外有關 PBL 的研究, 在學習理論的驗證研究十分缺乏; 由於 PBL 學習的最終成果不易證明, 因涉及太多變項及研究限制, 而且成果亦難以在短期內顯現, 故有關 PBL 在學習理論的驗證便十分重要。我們在希望獲得最終的成果之前, 必須要確認 PBL 是否真能達到它所希望達到的學習能力。

有關 PBL 學習效果的評估工具不多, 中文的、有跨文化驗證的更少, 所以相關的研究工具發展是進行研究的首要工作; 也可以提供日後其他類似研究使用。由於每個學校的 PBL 課程在內容設計與進行方式都有差異, 國內與國外的差異更大, 因此對於 PBL 的評估最好能做到國內與國外的比較, 以了解彼此間在學習效果上是否會有明顯的差異。基於上述原因, 國外現有的研究工具成為本研究的優先考慮。經過文獻找尋後, 我們發現 Sherbrooke 大學的 Tutotest 在問卷的理論架構、信效度表現與所評估的內容最符合本研究的需要。

在 PBL 的學習理論中, 知識在記憶層面的結構、學習者對知識的理解與批評能力是 PBL 的訓練重點。因此本研究利用概念圖嘗試評估學生學習 PBL 後, 知識的階層結構與理解知識的能力。概念圖(Concept mapping), 或稱認知圖(cognitive mapping)是由 Novak 等人所創(Novak 1972), 目的是瞭解學習者的思考結構與外顯知識(externalize knowledge)同時希望驗證 Ausbel's 的階級記憶理論(hierarchical memory theory), 瞭解學習者在學習新知識時是如何與舊知識作連結。

所謂概念圖是將某事件的重要概念用圖示來表示出來, Danseran 認為概念圖有三個特徵: 階級結構、連結和群聚。所謂階級結構是指下層概念是包含在上層之內, 連結代表某一概念會引致另一個概念, 群聚代表某些概念共有的特性。上述有關概念的性質都可以用概念圖表達出來, 概念圖的基本假設是人在學習新事物時, 對新事物所賦予的意義是受到原來的知識影響, 與原知識產生連結並獲得一個定位, 當我們繪畫概念圖的時候, 便會將概念與其間的關係表達出來。

概念圖常常用作一種輔助學習或教學的方法, 概念圖可以增進理解的學習, 幫助學習者整理大量的知識, 瞭解知識間之相互關係與整體性, 同時促進學習者進一步思考。概念架構是批判性思考與問題解決的基礎。亦有學者利用概念圖作為協助教學內容的編制。有關利用概念圖作為評估工具則較少。West 等人嘗試利用概念圖作為學習評估的工具, 對 21 位住院醫師做研究、測驗概念圖的計分方法的信、效度, 初步結果顯示其評分方式具有可接受的再測信度, 並可以顯示住院醫師接受指導前後概念的差異。

本研究目的如下:

- 一、針對台大醫學院 PBL 課程的實施狀況設計合適的學習評估方法。
- 二、根據 PBL 的學習理論評估目標學習行為的達成情況。
- 三、建立相關的評估工具與流程。
- 四、探討學習評估對學生整體成績的預測效度(predictive validity)及學生自評與教師評估的相關關係。
- 五、以概念圖作為學生理解能力評估工具的可行性及相關因素。
- 六、研究 PBL 是否影響學生理解能力。

研究方法與步驟

本研究以台大醫學系二至四年級學生及小班教學老師為 Tutotest 的研究對象，以及實習醫師為概念圖繪製及評估的研究對象。

一、PBL 學習效果的評估工具

Tutotest 為 Sherbrooke 大學為評估 PBL 所編製的工具(Hebert & Brave, 1996)，目的是評估學生在 PBL 中培養的態度與能力，理論架構是參考 Flanagan 的“Critical-incident”技巧，只選擇對於某些目標有關的可觀察到的事件與行為。原問卷之因素分析結果包括「PBL 學習技巧」、「團隊工作的效能」、「溝通領導能力」、「科學懷疑態度」與「同儕尊重」等五個向度。原問卷有 44 個題目，回答分為「很少或沒有」、「有時」、「常常」、「經常」四分評量，亦可選擇「無意見」。原問卷的因素分析依 eigenvalues 大於 1 的標準共獲得 4 個因素，共可解釋 82% 的變異量，內部一致性分析依因素 1 至 4 分別為 0.98、0.93、0.91 與 0.77，再測之相關係數為 0.46，Tutotest 結果與教師一般評分的相關為 0.64，與學生的筆試成績的相關為 0.39。

本研究將翻譯 Tutotest 成為中文，再反譯回英文，比較翻譯後的差異；並修改明顯不適用的地方編制成中文的 Tutotest^C 問卷。問卷經專家作文字修改與內容效度評估後開始進行前測。前測方法是以選取 50 名曾經接受 PBL 教學的實習醫師為對象，經研究者解釋研究目的與問卷填寫方法後施測，施測 2 週後再進行重測，結果經題目分析、內部一致性分析，再測信度與因素分析後作進一步修改成第一版中文 Tutotest^C 問卷。

二、中文版 Tutotest^C 的施測

以醫學系二、三、四年級學生為施測對象，台大醫學院醫學系二至四年級皆有 PBL 課程，醫二為 PBL 的熟悉階段，第一學期以人文社會議題為討論內容，第二學期開始加入生物化學的討論內容，醫三為解剖、免疫學、生理學等整合課程的 PBL 教學，醫四為病理、藥理與臨床醫學課程。PBL 課程以 7 至 8 人的小組進行，每學年重新編組，指導老師亦每年更換。在施測時為使學生能夠適應小組的運作，不致影響評估結果，因此選在第一學期結束前與第二學期期中各施測一次。

本研究以 Tutotest^C 施測結果除作個人與組間的分析外，並將 Tutotest 學生自評得分與老師評分結果作相關比較。

三、學生的認知結構與理解能力評估工具的編制

測驗流程的建立：參考文獻中概念圖的繪製方法，編制概念圖的繪製流程、說明與指導語，再選擇 24 位醫學系 5 年級學生為測試對象，選擇一個常見的疾病如頭痛，在上課時安排半小時時間讓學生繪製概念圖。依繪製結果評估學生是否瞭解正確的繪製方法。接著將同學分為三組進行焦點小組訪問，瞭解繪製概念圖的困難與學生的思考路線是否能在概念圖表現出來。依據焦點小組分析結果修改概念圖說明與指導語，並訂出適當的繪製時間，相關資料的提供等，制訂概念圖繪製的標準程序。

四、概念圖施測與評估

以在台大醫院家庭醫學部實習的六、七年級實習醫師為研究對象，由於台大醫

院之見習與實習醫師除了本醫學院學生外，同時接受其他學校學生作見習與實習訓練，由於部分醫學院的 PBL 課程開辦較晚，有很多六、七年級學生尚未接受過 PBL 課程，因此本研究將見習與實習醫師分為曾接受 PBL 教學與未曾接受 PBL 教學兩組，利用見習與實習醫師在家庭醫學科見習與實習時給予概念圖測驗。

施測方式依照第一年所發展之施測流程，施測結果由曾參與第一年概念圖評分研究的臨床老師負責評分，評分採取匿名方式，學生資料不會出現在測驗紙上。見習與實習醫師在本科的臨床表現成績，包括臨床知識、團隊合作、理解思考等向度的評分，依原有之評分方式評分，其結果除做為家庭醫學科學生成績外，並與概念圖評分結果比較兩者相關的程度。為了解評分的評估者信度，每個概念圖皆由兩位老師分別獨立評分。兩組老師的評分結果進行評分者間一致性分析(interrater reliability)，依據分析結果檢討評分方法。

結果

一、Tutotest^C 研究部分

小班老師共完成 370 份有效的 Tutotest^C 評估，在二、三、四年級的完成率分別是 92.7%、93.3%與 91.7%，三個年級的完成率並無統計上明顯的差異。三個班級的整體評量平均成績分別是二年級 86.0 ± 2.1 (77~92)；三年級 86.3 ± 2.2 (81~95)；與四年級 85.4 ± 3.6 (60~92)。各年級的得分分佈如表一。四年級之平均成績明顯低於其他兩個年級 ($F=4.05$, $d.f.=2$, $p=0.018$)。

Table 1. Distribution of Tutotest^C scores in the second, third and fourth year medical students

Scores	Second year	Third year	Fourth year	All students
0-19	9 (7.2)	2 (1.6)	2 (1.6)	13 (3.5)
20-39	39 (31.2)	33 (26.8)	41 (33.6)	113 (30.5)
40-59	43 (34.4)	56 (45.5)	59 (48.4)	158 (42.7)
60-79	24 (19.2)	23 (18.7)	13 (10.7)	60 (16.2)
≥ 80	10 (8.2)	9 (7.3)	7 (5.7)	26 (7.0)
Mean ± S.D.	48.7 ± 20.5	49.9 ± 17.9	46.4 ± 15.5	48.4 ± 18.0

Tutotest^C 與整體評量間的關係

整體評量成績與 Tutotest 成績有中等強度度相關($r=0.44$, $P<0.001$)，與 Tutotest^C 各分量尺亦有不等程度之相關：「團隊工作的效能」($r=0.42$, $P<0.001$)；「PBL 學習技巧」($r=0.42$, $P<0.001$)；「溝通領導能力」($r=0.35$, $P<0.001$)；「科學懷疑態度」($r=0.35$, $P<0.001$)；「同儕尊重」($r=0.15$, $P=0.003$)。

題目分析與信度檢定

Tutotest^C 各題目之平均得分介於 0.49 至 .09 之間，標準差介於 0.64 至 0.93 之間，skewness 介於 -0.58 至 1.44 之間，Item-total 相關介於 0.39 至 0.78。Tutotest^C 之內部一致性 Cronbach's α 係數為 .97，兩週之再測信度在 28 位同學之評估中，Pearsom 相關係數為 0.85。

因素分析

Tutotest^C 資料之 KMO 值為 0.952，Bartlett's 圓形檢定為有意義($X_{9462}=9824.08$, $P<0.001$)。顯示資料適合進行因素分析。因素抽取方式乃參考 Kaiser 條件。Cattell's 碎石圖與平衡分析 (Parallel analysis) 之結果。顯示適合之因素數目為四個因素，以主成分分析及 Varimax 轉軸共可解析 63.7% 之變異量。第一及第三因素包含所有的「PBL 學習技巧」題目與「科學懷疑態度」及 10 題「團隊工作的效能」題目。第二個因素包含 12 題「溝通領導」題目及 3 題「小

組討論效果」題目。「同儕尊重」則集中在第四個因素，四個因素之內部一致性測驗 Cronback's α 分別為 0.96、0.93、0.90 與 0.92 (表二)。

Table 2. Exploratory factor analysis of the Tutotest^C

	Factor 1	Factor 2	Factor 3	Factor 4
Factor 1				
Presents materials in an organized fashion during the synthesis	.836		.193	
Approaches the problem in a systematic, organized fashion	.796		.269	.115
Shares own information	.724	.179	.162	.261
Answers questions from colleagues	.724	.136	.189	.160
Presents own findings during the synthesis	.719	.176	.290	.254
Helps to keep the group task-oriented	.705	.181		.348
Uses appropriate scientific terminology	.704		.230	
Identifies and clarifies the elements of the problem	.679	.209	.350	.201
Participates actively in defining learning objectives	.662	.241	.303	.192
Presents to the group relevant complementary details arising from own investigations	.630	.304		.367
Asks questions to clarify obscure points, enhance understanding, or stimulate the discussion	.572	.438	.314	.178
Makes links with prior relevant readings, experience, or knowledge	.554	.354	.224	.207
Priorities the elements brought up during the discussion according to their importance	.539	.421	.247	.118
Demonstrates that he or she is up-to-date	.530	.421	.363	.212
Participates in each step of the problem	.522	.284	.442	.253
Contributes to the good mood of the group	.485	.384	-.169	.445
Summarizes what has been said, draws conclusions	.479	.354	.376	.397
Brings to the group information from references other than those suggested	.473	.439	.381	.210
Cronbach's $\alpha = 0.96$				
Factor 2				
Makes the group aware of time constraints (proposes or enforces realistic time limits)		.741	.251	
Is worried about the discomfort of another member of the group	.115	.730		.186
Encourages participation by others	.332	.694		.125
Verbalizes own well-being or discomfort in the group	.275	.689		
Indicates disagreement when a student or the tutor monopolizes the conversation	.111	.675	.269	-.194
Approaches different supplementary aspects of the problem (legal, ethical, humanitarian)	.203	.674	.293	.169
Indicates disagreement when the tutorial drags on to the point of becoming unproductive		.646	.243	-.191
Participates actively in the group evaluation	.277	.644	.308	.322
Brings the group back to the ongoing step	.426	.628	.221	.170
Respectfully points out errors made by others	.334	.620		.364
Breaks vicious circles when the group is going nowhere	.379	.595	.293	.216
Offers constructive criticism to the group and	.380	.570	.352	

colleagues				
Indicates disagreement or disapproval (without withdrawing or becoming silent or apathetic)		.561	.333	
Is introspective and tries to go further through questions addressed to the group	.412	.513	.393	.256
Shows evidence of a critical sense in regard to the presentations of other students	.234	.508	.467	-.174
Cronbach's $\alpha = 0.93$				
Factor 3				
Proposed explanatory hypotheses	.378	.242	.723	.131
Tests or critiques proposed hypotheses	.335	.339	.703	
Show evidence of a critical sense in regard to references and to the contradictions to which they give rise	.155	.407	.642	
Proposes an overall explanation or a schema unifying various hypotheses	.472	.330	.590	
Does not cite references but contributes to the synthesis of new knowledge	.430	.228	.556	
Intervenes appropriately in relation to the ongoing step of the problem	.425	.316	.550	.172
Formulates research questions following the study of the problem	.333	.479	.518	.161
Cronbach's $\alpha = 0.90$				
Factor 4				
Respects the right of all members to speak	.234		.111	.867
Communicates with others without hostility	.189			.863
Respects the values and opinions of colleagues	.223		.140	.851
Accepts criticism	.239			.830
Cronbach's $\alpha = 0.92$				

Principal component analysis with varimax rotation and Kaiser normalization.

Factor loadings of less than 0.1 were not shown in the table

學生自評與老師評分間的關係

學生與老師評分的平均值除了三年級外，二及四年級的學生自評都比老師評分為高。而老師整體性評分在四年級的平均值又較其他兩年級為高。Tutotest^C的學生自評與 Tutotest^C的老師評估，及 Tutotest^C的學生自評與老師整體性評分的相關均不高。Tutotest^C的老師評估與老師整體性評分則有中等強度的相關（表三）。

在 Tutotest^C的各分量尺中，學生自評與老師評分間的關係均很低。祇有三年級在「團隊工作的效能」、「PBL 學習技巧」、「溝通領導」三量尺中學生自評與老師評分有輕度相關（表四、五）。

Table 3. Agreement between students' self-assessment and tutors' assessment on students' performance in PBL tutorials

	S-Tutotest ^C (1)	T-Tutotest ^C (2)	Global score(3)	Correlation		
	Mean(SD)	Mean(SD)	Mean(SD)	(1)&(2)	(1)&(3)	(2)&(3)
Second year	48.8(14.4)	47.9(18.2)	86.0(1.9)	0.16	.25*	.44**
Third year	49.2(12.8)	49.9(16.6)	86.3(2.2)	0.28**	.40**	.47**
Forth year	49.8(13.0)	47.2(15.5)	85.8(2.7)	0.07	.25*	.37**
All students	49.3(13.4)	48.4(16.8)	86.0(2.3)	0.17**	.29**	.42**

P<0.05, **P<0.01

Table 4. Correlation between Tutotest^C subscales rated by students and tutors from grade 4 and all PBL students

	All PBL students									
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Eff(S)						.167**				
2. Abi(S)							.160**			
3. Com(S)								.192**		
4. Sci(S)									.083	
5. Res(S)										-.021
6. Eff(T)	.148									
7. Abi(T)		.162								
8. Com(T)			.228							
9. Sci(T)				.054						
10. Res(T)					.237					

Grade 2 students

Eff, *Abi*, *Com*, *Sci*, and *Res* represent the five subscales: effectiveness in group work, ability to master the PBL method, communication skills, scientific curiosity, and respect to others of the students' version (S) and tutors' version (T) of the Tutotest^C P<0.05, **P<0.01

Table 5. Correlation between Tutotest^C subscales rated by students and tutors from grade 2 and grade 3 students

	Grade 3 students									
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Eff(S)						.263**				
2. Abi(S)							.255**			
3. Com(S)								.271**		
4. Sci(S)									.132	
5. Res(S)										-.137
6. Eff(T)	.085									
7. Abi(T)		.055								
8. Com(T)			.077							
9. Sci(T)				.049						
10. Res(T)					-.153					

Grade 4 students

Eff, *Abi*, *Com*, *Sci*, and *Res* represent the five subscales: effectiveness in group work, ability to master the PBL method, communication skills, scientific curiosity, and respect to others of the students' version (S) and tutors' version (T) of the Tutotest^C
P<0.05, **P<0.01

二、概念圖研究部分

概念圖製作與評估流程

概念圖依不同的目的而有多種不同的繪製方式。一般可依繪圖之限制方式分為填充法 (fill-in methods) 與自由繪圖 (free construction) 兩大類 (表六)。

Table 6. Methods of concept map production

Mapping methods	Descriptions
Fill-in the map	The structure of the map was provided Students are asked to fill-in the missing parts
Fill-in nodes	The nodes (concepts) are missing
Fill-in links	The links (relation between concepts) are missing
Construct a map	The structure of the map not provided and students are asked to construct their own map
Concepts provided & structure suggest	The concepts and linking words are provided to construct the map

Concepts provided	Only concepts are provided
Free construction	No concepts and linking words are provided

由於本研究的研究目的乃為瞭解學生對某一概念的瞭解程度，因此採用自由繪圖方法較不會造成學生的限制，同時較能顯示學生間能力的個別差異。由於繪圖會受到知識內容的影響，為評估學生的能力故決定繪圖時不可以參考任何書本與資料。

概念圖的評分方式

由文獻中整理發現概念圖有六種主要的評分方式。由於本研究的重點是比較學生間的差異，故不採用標準圖（master map）作比較，故本研究只採用其餘四種評估方式評估（表七）。

Table 7. Scoring concept maps

Score components of maps		
Structural methods		
Novak & Gowin (1984)	Valid relationship	1
	Hierarchy	4
	Cross-link	10
	Example	1
Markham et al (1994)	Concepts	1
	Valid relationship	1
	Branches	1
	Success-branches	3
	Hierarchy	5
	Cross-link	10
Austin & Shore (1995)	Example	1
	Causal link	3
	Correct link	2
	Possible link	1
Errors	Errors	0
	Holistic method	
	Overall judgment of the map rated on a scale from 1 to 10	

概念圖的評分結果

本研究共獲得 33 份以頭痛為主問題作分析之概念圖。其中 15 份來自醫六學生，18 份來自醫七學生。有六位學生是台大以外的醫學院校。學生都可以在 1 小時內繪製完成。使用無母數分析(Mann-Whitney Test)顯示概念圖得分與學生年級及學生是否有 PBL 學習經驗無關。表八為四種評分的描述分析，顯示老師評分具有極大的差異，尤以用 Novak & Gowin 與 Markham 評分法之差異最大。

Table 8. Descriptive Statistics of concept map scores (66 ratings)

Structural methods	Mean (S.D.)	Range
Novak & Gowin method	54.97 (50.92)	8.00 - 238.00
Markham method	124.34 (107.46)	12.00 - 592.00
Holistic method	5.05 (1.81)	1.00 - 9.00
Relational method	53.36 (39.10)	7.00 - 195.00

以 Pearson 相關作評估者間信度分析，結果顯示 Novak & Gowin 法缺乏評分者信度($r = -0.08$)，Markham 法祇有輕度之信度($r = 0.38, P < 0.05$)，而整體法($r = 0.57, P < 0.01$)與關係法($r = 0.64, P < 0.01$)則具有較佳之評分者間信度。以組內相關分析法(intraclass correlation)分析四種評分法之評分者信度，亦發現與相關分析類似的結果（表九）。

Table 9. Intraclass correlation coefficients of the four scoring methods

	Novak & Gowin	Markham	Holistic	Relational
Single measure intraclass correlation	-0.18	0.005	0.50	0.49
Average measure intraclass correlation	-0.49	0.01	0.67	0.62
ANOVA	F=12.14 P=0.002	F=16.99 P=0.0003	F=2.64 P=0.114	F=4.96 P=0.033

利用 Pearson 相關分析顯示上述四種概念圖的評分皆與學生的實習分數無統計上有意義的相關。

討論

本研究發現中文版 Tutotest^c 具有良好的信效度，由於本研究的學生樣本數比發展原問卷的研究為多，故因素分析的結果應更為穩定。

由於台大醫學院與其他亞洲大部分的醫學院校一樣，採用小班與大班混合教學課程。而且小班以問題為導向的教學課程(small group PBL)，在設計上與加拿大 McMaster 大學不盡相同，故由歐美國家所發展的評估工具在台大醫學院的適用性有待研究。本研究已初步證明中文版 Tutotest^c 在本院小班教學課程評估具有良好的信效度。但小班老師使用本問卷後，普遍認為 44 題題目稍嫌過多，希望能盡量簡短，以減輕評估的負擔。因此將來希望在不影響問卷信效度的情況下縮減題數。

做為一個良好的評估工具，中文版 Tutotest^c 尚需要證明有能力測量學生能力的改變，這方面有待將來長期性的追蹤研究來澄清。

小班問題為導向的學習的主要目標，是培養學生終身自我學習的態度與能力，而自我評估的能力是作自我學習的基本要素。學生必須有能力自我反省，瞭解學習的優劣與不足，才能進一步自我學習。不幸的，本研究與大部分國外文獻報告一樣，發現學生自評與老師評估間存在有重大差距。老師與學生在學習能力表現的不同意見，除了可能反映學生缺乏自我評估能力之外，老師評估的偏差亦是必須關注的問題。過去很多研究指出評估者會有以偏蓋全等評估的問題，必須進一步研究澄清。本研究發現老師對學生之整體性評估與中文版 Tutotest^c 評估之分數雖有相關，但相關程度不夠高，顯示兩種評分方式所針對的學習能力重點不同，而整體性評估更容易有所謂的印象分數，容易落入以偏蓋全的陷阱。

在概念圖研究部分，本研究發現實習醫師間對於一個常見的健康問題—頭痛的理解存在十分大的差異，這個差異可以由他們所繪製的概念圖的結構複雜程度與老師評分的高低差距顯示出來，這種能力差距並不容易由一般的臨床觀察中獲得。

有關概念圖的評分，學者提出很多不同的評分方式，概念圖之原作者 Novak 等人之評分方法及其後一些學者的修改版本，其評分過程都較為複雜，而對於不同的理解能力又有不同的加權計分方式，本研究發現採用 Novak 或 Markham 之方法，都無法獲得良好的評估者間一致性，與國外研究的結果具有良好的評估者間一致性大不相同。其原因一方面可能是評估者訓練不足，另一方面國外研究多用填充法繪圖，與本研究之自由繪圖方式不同，是否因繪圖方式不同導致評估者一致性降低尚須進一步研究。

學生概念圖的繪製能力與六或七年級，及是否有小班問題為導向的學習經驗無統計上相關，顯示一年的臨床學習差距並未明顯增加對臨床問題的理解能力，採用小班與大班混合教學課程可能並無增加對臨床問題的理解能力。由於本研究的個案數太小，加上外院校到本院實習的學生本來就有一個成績門檻。很可能會產生研究結果的偏差，有待搜集更多的概念圖及作學業成績的調整來進一步分析。

參考資料

1. Neufeld VR, Woodward CA, MacLeod SM. The McMaster MD program: a case study of renewal in medical education. *Academic Medicine* 1989;**64**(8):423–32.
2. Khoo HE. Implementation of problem-based learning in Asian medical schools and students' perceptions of their experience. *Medical Education* 2003;**37**(5):401-9.
3. Leung KK, Lee MB, Lue BH, Yang PM, Hsieh BS. The implementation of problem-based learning at National Taiwan University College of Medicine. *Journal of Medical Education (Taiwan)* 2001;**5**: 273-80.
4. Hébert R, Bravo G. Development and validation of an evaluation instrument for medical students in tutorials. *Academic Medicine* 1996;**71**:488-94.
5. Des Marchais JE, Vu NV. Development and evaluating the student assessment system in the preclinical problem-based curriculum at Sherbrooke. *Academic Medicine* 1996;**71**:274-83.
6. Nunnally JC, Bernstein IH. *Psychometric Theory*. New York: McGraw-Hill, 1994.
7. Bryant FB, Yarnold PR. Principle components analysis and exploratory and confirmatory factor analysis. In: Grimm JG, Yarnold PR (eds). *Reading and understanding multivariate statistics*. American Psychological Association Books, 1995.
8. Astin LB, Shore BM. Using concept mapping for assessment in physics. *Physics Education*. 1995;30:41-5.
9. Ausubel DP, Novak JD, Hanesian H. *Educational psychology: a cognitive view* (2nd ed.). New York: Werbel and Peck.
10. Barenholz H, Tamir P. A comprehensive use of concept mapping in design, instruction and assessment. *Research in Science and Technological Education*. 1992;10:37-52.
11. Briscoe C, LaMaster SU. Meaningful learning in college biology through concept mapping. *The American Biology Teacher*. 1991;53:214-9.
12. Bruner J. *Acts of meaning*. Cambridge, MA: Harvard University Press, 1990.
13. Hegarty-Hazel E, Prosser M. Relationship between students conceptual knowledge and study strategies. Part 1: Student learning in physics. *International Journals of Science Education*. 1991;13:303-12.
14. Holley CD, Dansereau DF. Networking: the technique and the empirical evidence. In Holley CD & Dansereau DF (Eds.) *Spatial learning strategy: techniques, applications and related issues*. New York: Academic, 1984: 3-19.
15. Markham KM, Mintzes JJ. The concept map as a research and evaluation tool: further evidence of validity. *Journal of Research in Science Teaching* 1994;31:91-101.
16. Marriam SB, Caffarella RS. *Learning in adulthood* San Francisco: Jossey-Bass, 1999.
17. Piaget J. *The psychology of intelligence*. Totowa, NJ: Littlefield, Adams, 1966.

18. Novak JD, Gowin DB. Learning how to learn. New York and Cambridge, UK: Cambridge University Press, 1984.
19. Novak JD. Concept maps and Vee diagrams: two metacognitive tools for science and mathematics education, *Instructional Science* 1990;19:29-52.
20. Novak JD. Learning, creating and using knowledge: Concept Maps™ as facilitative tools in schools and corporations, Mahwah, NJ: Lawrence Erlbaum Associates, 1998.
21. Rowell R. Concept mapping: evaluation of children's science concepts following audio-tutorial instruction. Unpublished doctoral dissertation, Cornell University, 1978.
22. Pankratius W. Building an organized knowledge base: concept mapping and achievement in secondary school physics. *Journal of Research in Science Teaching*. 1990;27:315-33.
23. Schmid RF, Telaro G. Concept mapping as an instructional strategy for high school biology. *Journal of Educational Research*. 1990;84:78-85.
24. Stice CF, Alvarez MC. Hierarchical concept mapping in the early grades. *Childhood Education*. 1987;64:86-96.
25. Surber JR, Smith PL. Testing for misunderstanding. *Educational Psychologist*. 1981;16:163-74.
26. Trowbridge JE, Wandersee JH. Identifying critical junctures in learning in a college course on evolution. *Journal of Research in Science Teaching*. 1994;31:459-73.
27. White RT, Gunstone R. Probing understanding. New York: Falmer Press, 1992.
28. Willerman M, Mac Harg RA. The concept map as an advance organizer. *Journal of Research in Science Teaching*. 1991;28:705-11.