

對智慧型個別教學系統 研究的批判分析

朱玲玲 譯

台北市立中正國中

這是一篇關於“智慧型個別教學系統”(Intelligent Tutoring System簡稱ITS)的評論。ITS是一個企圖模仿人類教學的最新一代電腦化教學系統。研究人員試着結合比以往那些系統更多的功能，它能夠真正分析學生的反應，而不只是判定學生答對題目與否，並且，當學生答錯時，能夠針對他的答案有所回應。

這和其他所謂的電腦化教學系統，有着極大的不同，那些已經問世將近二十年，但一向不為老師們所接受的系統，只是由電腦提出一連串設計好的問題，以及為了糾正學生前一題的錯誤，所選擇的一系列題目而已。

到目前為止，ITS仍處於研究階段。筆者的評論是根據翻閱電腦及教育方面的文獻。儘管ITS和教育之間有明顯的相關性，將其結合運用的遠景也非比尋常。然而有關於ITS的論文却很少探討教育。綜合這些閱讀的心得，筆者認為ITS的工作有兩個在方法上主要的缺陷。

第一，ITS沒有建立完善的學習模式，它着重於高科技的使用而非教育上的要求。許多系統往往避開學習模式這個關鍵性的問題，有些系統只想讓你清楚地明白某些學習理論，而不提供這些理論有效的證據。於是它們幾乎採用同一種層次分明的，由上而下的，目標導向的循序漸近的學習模式。而這種大多數研究者視為當然，即使不是唯一也是最好的學習模式，卻是所有教育學家和社會科學學者所反對的方法。

第二，肯定ITS的聲言，所根據的測試是典型的控制不良和報導不完整，結論不能讓人信服，有的甚至於完全缺乏測試。

縱使有模式不妥和測試結果的問題，一般的文獻中對所有電腦輔助教學系統，都有着很高的評價。

一、模 式

一個具備了完整的教與學基本模式的電腦化教學系統，可以視為對教育以及人工智慧 (Artificial Intelligence 簡稱 AI) 的真正貢獻。要獲得這樣的模式談何容易，由教育學上的教學理論來印證 ITS 的使用，對於什麼是最佳教育技巧，教育專家們就很少意見一致的了。這也許意味着教育是“專家系統”(expert system 試着複製人類專家技術的電腦系統)裏很糟糕的部份。雖然如此，ITS 仍在實現中，接着我們將討論這其中細節問題。

原案分析是建立教學模式的必要條件。所謂原案分析是指，收集真實的老師與學生教學狀況的資料，並且對可能具有形成有結合教育模式的部份加以分析。社會科學建立了許多很好的收集資料方式，包括定性的測驗和調查，以及定量的研究。

從事 ITS 的研究人員採用定量方法為主，這似乎犯了一個錯誤，在學習過程中，不可捉摸的人際關係扮演了很重要的角色，例如說，我們都會為了自己所喜愛的老師更用功學習，並且也會學得更好，沒有敏銳的定性資料收集方式，就會遺漏這種不可思議的情況。這種原案分析的工作，必須在學生和資深老師上，投注很多時間，舉例來說，不花時間去觀察高中上代數的班級，不花時間和高中的老師及學生討論代數教學，就無法發展出一套完善的高中代數教學的模式。這樣費時而困難的資料收集和分析工作，電腦科學家們是很少有經驗的。

有些 ITS 研究人員根本沒有作好這項原案分析的工作。有個研究指出它的兩個系統，建立於豐富的資料來源，而這些資料只從三個到四個學生身上收集而來；另一個研究指出它分析了十二個學生，試想這種結果可以說服一位在教室中執教的老師嗎？有些論文甚至於沒有提到原案分析這回事，研究者沒有去作，更不覺得有必要在他的研究報告中提出。

有兩篇論文作了原案分析的強調，研究者在論文中企圖舉出在教學原則概念上的錯誤，並想發展出一套在主要領域中統一的基本模式。和其他研究所不同的，在於它致力於原案分析的程度，超過系統的設計與實施，它相信在恰當地考慮其電腦化之前，更多收集資料的基礎性工作是非常必要的。這些研究所建立的模式太過於高層次而抽象了，在決定系統是否能夠在電腦上實施之前，有更多基本而詳細的模式需要發展和建立。不經過正確的原案分析，ITS 往往遵行一種試行錯誤的程序，先產生一個版本，然後測試，根據測試修改出新的版本，再經測試，直到完成。在本文下一部份會討論有關正確

測試的細節。有一篇論文，甚至於只經過一個生產與測試的循環就發表了，而且還揣測系統若如何改進，結果將會如何。

“錯誤範例的一覽表”並不是一種模式。

既然沒有教學的標準模式，許多系統便以學生學習該特殊技巧所犯的錯誤的集合來取代，舉例來說，減法學習的程式，就以 110 種基本常見的缺點和 58 種補充說明組成混合交錯的格子，以及學習 LISP 語言的教學程式就例舉了 325 個規則和 475 個易犯錯的法則。也許分析對與錯的規則 ITS 中很有用的部份，但是錯誤類型的一覽表只是建立教學模式的原始資料而已，不經過分析是無法闡明教學的程序和步驟。再說，要求列出所有的錯誤類型會把 ITS 限制在很簡單的範疇。

Matz 列證了一個不同的研究，我相信是有展望的研究，她不再列舉所有錯誤類型，而代之以審定學生問題的資料，和進行一個能說明且誘發錯誤（也可能是正確答案）的高層次概念的系統。這個強調原案分析以及用學生學習狀況為基礎而建立模式的工作是非常精彩的，她寧願從事這些資料收集、分析和模式建立等基礎工作，而不是投入在系統的實作。

“資料表徵”並不是一種模式。

知識領域的表徵是 ITS 工作的主要重點，有個研究認為表徵資料和建立師生教學模式一樣重要，也就是說他認為 ITS 的科學技術和內容一樣重要。資料表徵是電腦科技重要的部份，但是對教學來說則不然，大多數情況，特別的表徵是用來對付技術上的問題，而非教學問題。如果教學系統是首要的，那麼資料表徵也僅只次之；因此，另外有個系統採用特別由教學期間所完成的資料來定資料表徵的發展方向，而不是一開始就隨意解釋資料，以與研究領域不相干的理由，選擇特別的資料表徵。

如果某種資料表徵有着認知理論上的基礎，硬要使用它似乎是合理的，可是幾乎所有的研究者並不暗示他們的資料表徵有此基礎。有一份研究強調了資料表徵的認知基礎，作者對使用被稱為“製作法則”的表徵不作任何的辯證，他說：「要問什麼是製造系統的證據，是很有趣的。製作系統的假說太抽象了，以致於無法落實在直接的實驗測試。“製作法則”是個好構想，因為許多 ITS 都採用它」。無論如何，因為通常對不同的資料表徵很少去分析，使得問題很容易地永遠存在一個又一個的系統中。

“實例”不是一種模式。

有許多重要的設計，完全取決於機械效益等技術因素的考慮，而忽略對教學的效益。例如，“DEBUGGY”（除錯系統）採用“全部問題證據”的流程，來決定一些已

知錯誤的子集合，這子集合再變成原有假設的集合。這流程並不具備教室中實際教學經驗，或者認知理論的基礎，只不過是它在電腦上執行很有效而已。

LISP 系統採用選擇題，學生可以從目錄中選擇教授者所提出問題的答案，並不是因為選擇題為最好的方法，而是因為選擇題可以改善由原始語文輸入以及要電腦來剖析原始語文的困難。雖然設計 LISP 教學系統的人，聲稱他的系統具備認知理論的基礎（當然是他自己的理論），但是他這種選擇預定答案的方法，是沒有任何理論根據的。他另外有關於選擇目錄的有效力的敘述完全不清楚。

這樣的實行理由會以無法預料的途徑影響教學程序，而有些是不良的影響。在實驗階段中，早些實現系統的壓力是必然存在的，卻不見得迎合教學者的興趣。「大部份教學技術家只注意 ICAI (Intelligent Computer-Aided Instruction) 的可行性，而忽視了 ICAI 的理論基礎何在」。

整體來說，除了根據實行者的直覺以外，很少有根據其他任何理論的證據；沒有教師與學生在教室中確認 ITS 實際經驗為主的有力證據；沒有真正專家（老師與學生）參與驗證這些系統的嘗試；參考認知心理學與教育學理論的作業也很少，而這些理論可能提出更實在的理由來評判系統的優劣。（有兩篇尚未實施的系統，在學習模式的建立上，作了更多參考的工作），結果是一堆直覺在互相競爭，不可能做實質上的評斷。

一份電腦化教學 (Computer-based Instruction 簡稱 CBI) 的調查，研究得到下面的結論「雖然 CBI 似乎經常都是無理論的，但是明顯地，CBI 具有廣大的研究基礎……」，我相信它似乎經常都是理論的，因為它的確如此，而我擔心 ITS 承襲了那些被教育家們批評為不切實際的教學軟體的傳統，以及似乎由那些不甚了解教室裏實際狀況的人們來設計軟體，而這些軟體將要使用在教室中。

此外，ITS 工作單純地強調技術，忽視其他那些每一個老師都知道對優良教學很重要的因素。我們不確知教師們是否將所有錯誤毫不遺漏地加以系統化（即使他們確實注意到常犯的錯誤類型），然而我們確知教師們考慮到許多錯誤以外的其他因素。例如，老師們明白要成功的教會一些東西，用角色模式舉例或滿足學生的注意和關懷等人與人之交互作用，和糾正錯誤一樣重要。電腦系統只能對付部分的教育問題，我們的實際經驗告訴我們，師生關係的建立等非技術性，非定量的因素是非常重要的，但是這些因素電腦可以運算嗎？

二、學習的型態

雖然已有的 ITS 系統才剛開始要明白地提出教學過程的模式，某些建立模式的途徑也明顯地不夠嚴謹，而所有我所看過的系統幾乎採用同一種學習的型態，那是一種層次分明的、結構化的、循序漸進的，由上而下而且目標導向的型態。即使有作者提出最佳的互動型態可以廣泛因應個別差異，在系統中仍然提出那種由上而下、由左而右的解答，而且誠訓學生們堅持逐步細節化的過程。LISP 的教授者理想是層次分明的、結構化的、由上而下的，而且教學目標層次分明。有一篇論文綜合所有 ITS 工作這麼說：「CAI 課程內容 (courseware) 因為提供個別照顧的功能而受到讚美，縱然如此，大多數 CAI 課程內容，對所有學習者有着一成不變的回饋與增強」。

ITS 的研究者隱約地假定只有一種學習型態是正確的或最好的，相反地，電腦科學領域以外的研究指出，有好多種成功的學習型態。Sherry Turkli 對兒童使用電腦的研究就是個例子，她說明即使像寫電腦程式，這種一向被認為是強迫而刻板式的學習，也有許多不同的學習型態，她描述了兩種型態的精通熟練——“軟的”和“硬的”，“硬的”精通強調預測控制、結構、目標和控制，“軟的”精通則強調互動，相互關係和調適。她的研究擁有廣博的原案分析，以及其他無論定性或定量的資料收集與分析的經驗，而且她的原案分析報告也比其他 ITS 系統完整，她更結合了人類學及心理學上的相關發現。

我曾讀過一個系統提出對學習兩個不同的見解：(1)劇本式觀點是階層分明的，循序漸進的，強調互不相關的事實。(2)功能式的觀點是非線性而相互影響式的功能型態，它着重事件的基本原因和相互關連。劇本式觀點符合了其他系統主張的看法，當研究者發現這種觀點不足以解釋最簡單的教學式交談的變化，他們便下結論說附加一些看法觀點也許有用。的確，結合上述兩種見解，可以說明半數以上教學的交談內容。

承認教學方式應該不只一種是由以下兩個重點來支持的，其一就是上面說過的，學習類型不止一種型態，其二便是衆所皆知的事實，教授不同知識，針對不同的學生要使用不同的解說方式。強迫學生接受唯一型態的教學系統，必然遺忘那些天生具有不同特質的學生。而對電腦系統一向自許為幾乎沒有限制、富彈性，而且適用於教育上的說法，是一大諷刺。

“是的，電腦具有特殊而不受限制的分支能力，而我們無法預見少許學習者一般可能的反應，是沒有關係的。在限定的範圍中，學習者就注定要回答那些程式所設定好的

問題”。

有位評論家提到“大多數 AI 系統有能力找出「一個」解決問題的途徑”。可是，在教學上，一種途徑是不夠的。

三、測 試

另一個 ITS 在方法上的缺陷就是測試。在所有我看過的文獻中（也是電腦科學家所提出的所有論文），測試資料通常支持下列論點：電腦會節省學習時間，電腦可以改善學生對學科的態度，電腦化教學增進學生學習能力，和傳統教學方式一樣好，甚至於更好。（由電腦和教師配合進行教學的補助性的 CAI 系統評價最高）更仔細審視這些包含教育學的文獻資料，會發現在測試中一些令人驚訝的缺失。

大部份 ITS 研究的結論，是由不正確的測試資料推導而成，這些測試資料包括太少的受試者和測試主題。教授 LISP 系統，幾何學教學系統和基礎代數課程的 LMS 系統，都只提及一次測試。幾何學教學的測試只有三個學生參與，而 LMS 系統測試只有六個學生。一份關於程式設計教學的論文報告“它似乎只能提供一系列同屬性簡單的程式設計，而有結合力的架構組織可以擴充到分析更複雜的程式設計工作”，但是，作者並沒提出更進一步的測試報告。

在某些案例中，只是由非正式的測試來作出結論，經過三個學生使用幾何學教學系統之後，作者說道：「我們相信學生比用傳統教法學得更快，但是我們無從考證這個信念」。

還有一些案例，由不確實的資料描述出正面評價的結果。有一篇論文只安排一個象徵性的談話來描述有關教學互動性的理論，而其中有 44 % 的意見交換並沒有被理論考慮到。那個只有六個學生參與測試所寫成的 LMS 系統研究報告中，有 56 % 的學生錯誤，該系統診斷不正確。作者提議增加一些額外的規則，宣稱可以讓系統有能力正確地診斷出全數 70 % 的錯誤，但是這種假說並沒有經過測試。他建議其餘所有的錯誤（只有一個例外）都可以分類成執行的和隨機的，然後將其忽略！他的結論說他的系統可以診斷所有的錯誤。

許多有關 CAI 的正面性評價，在教育學的文獻中提出，我曾看過兩篇研究 CAI 效度的“變異數分析”的論文，一篇審視 59 種研究，另一篇有 51 種研究，而兩篇均由同一個人所撰寫。沒有一篇論文考慮這些獨立不相關的研究是否可以有效地比較。兩篇論

文都得到有利於CAI的結論，CAI得到比較高的成就積分，比較低的退步比率，學生對用CAI教學有比較大的喜好。以下是其中代表性的報告：

“59個研究中，有54個研究在成就評量中具有有效的統計資料，54個研究中有13個研究的結果對CAI有利，只有一個顯示傳統教學比較好，因此，CAI獲得優良的成就評量結果”。

另一種方式陳述相同的資料：

“59種研究中，22%有利於CAI，2%有利於傳統式教學，以及76%並不偏向任何一種方式或是未達統計上的顯著水準，結論是無從證明CAI具有優良的成就評量結果”。

這裡有更多上述兩篇論文所根據之基本資料：

- 成就評量：在第二篇論文中，48%偏向CAI，4%偏向傳統式教學，而76%兩者均不。
- 退步比率：（只有一篇論文有此報告）3%有利於CAI，2%有利於傳統式教學，95%兩者皆不或是無顯著差異。
- 學生態度：其中一篇，有3%喜歡CAI教學，97%都不喜歡或是在統計上無顯著差異；另一篇中，有20%喜歡CAI教學，80%都不喜歡。
- 教學時間：在其中一篇論文裏，選擇51種研究中的兩種，而兩種的結果均有利於CAI，另一篇論文則挑選59種研究中的8種，結果8種均有利於CAI。

有份研究明白選擇不同的教師會影響教學結果，當同一個老師同時採用CAI教學，以及傳統方式教學，測試結果顯然偏向CAI。雖然如此，作者僅只對CAI作正面效應的聲言。

當電腦在電腦上應用顯示出缺乏嚴格的基礎的研究特性，都輕描淡寫地帶過。

另外一些系統的測試，被不適當的實驗設計，非隨機分配主題以及不正確的假說來歸類。這些不正確的假說令不同的族群有相同的結構和實驗者心理傾向，有過度簡短的測試和可疑的分析技術；簡而言之，很少評價是有事實根據的。“因此，最近有些陳述CAI為有教育性和有效性的，是完全沒有科學根據的結論”。（這篇報告同時也指出，類似這種缺乏證據的情況，也發生在其他，非電腦的教學方法評估上）。

這兩篇有關CAI效度的變異數分析之研究，都提到像同儕影響的教學技巧等方法，比CAI更有效，這些對一個成功的教學很有幫助的方法，並沒有被ITS的設計者注重和研究。

四、結 論

ITS 的技術，具有兩個主要的方法上缺陷：

1. 該系統並沒有落實於具體的教學模式。原案分析應該在建立模式之前施行，但是原案分析作得很少，而且根本沒作定量的分析。ITS 模式應該經由那些將要使用 ITS 系統的教師和學生們確認，而研究者似乎根本不去請教他們。

2. 測試不完備而且不確實，有些根本沒有測試。電腦化教學的資料被混淆。在教育文獻上普遍可見，對 ITS 以及其他電腦化教學系統的正向評價，都以一些有缺陷的測試結果為基礎。

所以，要知道如何選擇在什麼 ITS 上用什麼解題流程計劃，是非常困難的。例如，Matz 和 Sleeman 先後用不同的技術來完成代數教學的計劃，Sleeman 參考了 Matz 先前的成果，而他未提及為何採用新的技術。在另外一個領域，Goldstein 和 Burton 分別提出兩個完全不同的解題方面的方案，這些不同的方案都具有某些直覺的部份，可是沒有一個具有鞏固基礎與測試足以令人信服，那麼讀者只好在研究者的直覺組合中選擇那個比較好。

不管這些，許多非常樂觀的電腦輔助教育系統的工作者，宣稱因為這些系統，教育將發生迫切的革命性的改變。一份 1980 年 CAI 的調查預測電腦化學習的成本效益還差強人意。1983 年的調查指出“許多研究在這方面努力討論，無疑地對教學實施有着意義深遠的影響”。（但是，它在另一處卻說“不管 ICAI 的承諾如何，沒有任何一個研究造成教學實施真正的衝擊”）。一位 ITS 的研究者預言“從電腦程式設計到英文作文，任何一方面的知識傳授態度將徹底的改變”。但是沒有教育者的接納，是不會發生徹底改變的。而這樣的接納似乎並不存在。在他發表這個說法的那一年，有個針對 4000 個人上的 CAI 系統之研究發表，在各相關領域中的教學者，有 96% 到 97% 的人並不接納 CAI。

不論 ITS 的潛能有多少，到現在為止所完成的系統，不論在科學價值或可能引起的教育型態轉變，還不如所受到之正面評價。一個針對 20 年來電腦背景的調查顯示，電腦在科學技術方面的確有所進步，但在教育方面則不然。

“20 年前，一套本質上和現在相同的電腦化教學話題，被討論過，事實上，我們仍然要面臨那些已經出現過的問題，而且也許比以前更多。但是，現在討論的層面卻被技術性地修飾過”。

ITS工作尚在未成熟的階段，以我讀過關於 ITS 的論文作基礎，以及用我和教師中老師、學生面談紀錄，我曾製作有關教學的電腦軟體，我建議 ITS 應該由下列過程得到助益：

- 在實施其他系統之前，和老師、學生們一起，花更多的時間在實際教學狀況上。
- 在發表充滿不合實情與錯誤結論的報告之前，要有更好的測試與結果分析。
- 在急迫的教育“改革”製造更多預言之前，要對科技在教育中的角色扮演更謹慎。

最後，ITS 應該要記住前述那些教育技術方面調查的結果：

“優良教師和良好課程內容發展者的貢獻、支持和能力，才能促成 CBI 等系統的潛在利益”。

當然，如果有足夠的，能奉獻而持續性高又有能力的老師和課程內容發展者，而且社會也保證支持他們，那麼誰會對 ITS 沒有興趣呢？！

參考資料

Ronni Rosenberg, "A Critical Analysis of Research on Intelligent Tutoring Systems". Educational Technology, November 1987, 27(11), P.25.