科學教育學刊
2001, 第九卷第三期, 299-322

Chinese Journal of Science Education
2001, 9(3), 299-322

如何能發現隱藏的規則? 從科學資優生表現的特色，探索提升規則發現能力的方法

林緯倫　連韻文
國立台灣大學 心理學系

(投稿日期：民國90年4月19日，修訂日期：90年8月22日，接受日期：90年11月1日)

摘要：形成假設，檢驗假設從而發現規則是人類進行分類、區辨、形成概念、建立因果關係等思考活動時所需具備的能力，也是科學研究的必備能力。本研究採用心理學中常用的規則發現作業－「246作業」，以實驗的方式，探討假設產生與檢驗的能力，並分析具有科學潛力的人才在此一認知能力上的特色。我們比較一般大學生與通過教育部甄選進入大學就讀的科學資優生在此作業的表現，如所預期，後者的正確率的確較高。分析其策略，他們在測試法上較一般生多採用「雙假設測試法」，而非過去研究所關注的反例法。此外，如所預測，在修正假設的過程中較一般生產生更多新角度的假設。這和我們過去發現一般大學生中成功者的特色相仿。我們討論採用雙假設測試法的可能優勢並考慮工作記憶資源的限制，據此我們指導另一組大學生採用一種「有限的雙假設測試策略」，結果一般大學生在假設檢驗推理作業上的成功率由26.7%大幅提升至81.8%。

關鍵詞：工作記憶、另有假設、科學資優生、假設檢驗、概念改變。

假設檢驗推理能力的重要性

在日常生活中，人們不斷地對現象作出解釋，再藉由新資料或新經驗的回饋來修正假設，以建構出我們對周遭生活環境的理解，對未來作更有效的預測。這種形成假設與修正假設的推理過程常是自動自發的，舉凡分類（金牛座的人通常有某某的性格）、區辨（是否為腸病毒感染?）、概念形成（「炫」的定義）、因果關係與解釋的建立（股市大跌的原因）都牽涉這樣的過程，藉由這樣的歸納歷程以發現一些生活中的規律和運行的規則，而對現象有更好的解釋與預測。在科學研究的領域中，也同樣在進行這樣的工作。科學家試圖瞭解自然界運作的規則，他們從已知中形成假設，然後尋求證據或反證來修正他們的假設。科學的發展除了基礎知識的傳承之外，更重要的是新知識的發現，此乃涉及了創造的歷程。以拖浪斯創造性思考測驗(Torrance Tests of Creative Thinking)聞名的心理學家Torrance（1989）

曾說過：「創造性思考是一種歷程，包括覺知到困難、訊息的不足、要素的缺失，從而猜測、形成假設，並評估與測試假設，儘可能修正假設並重新驗證，最後將結果傳播出去」。他的看法點出了新假設的產生與評估測試是創造過程中的重要環節。一個有創造力的科學家，需要能從現象中形成適當的假設（規則），並且有好的策略去檢驗其假設，而後能整合新資料來修正原有假設，或看出嶄新的規律性。例如發現苯環結構的化學家 Kekule 在線性結構無法解釋實驗現象時，能突破原有的架構，利用環狀連結方式來重新解釋現象，因而成功地解決了難題。假設檢驗推理歷程的研究將可使我們更瞭解這種富創造性的思考過程，這對科學教育具有重要意義。

此外，科學教育的一個重要目標即是要診斷出學生已有的迷思概念，繼而進行「概念改變」，讓學生學習到正確的科學知識（e.g., 宋志雄、林曦和徐順益, 1993; 洪振方, 1996; 郭重吉和吳武雄, 1989; 張川木, 1995, 1996; 謝志仁和郭重吉, 1993; Chang, 1993）。學生藉由新訊息的獲得，以對自己的信念或假設進行修正或改變，其所涉及的歷程即為假設檢驗與假設修正的推理過程。透過對於此一歷程的研究，也可以幫助我們瞭解人們如何檢證概念與改變概念，因而對於如何幫助學生修正概念、避免陷入既有迷思概念有所啟發。

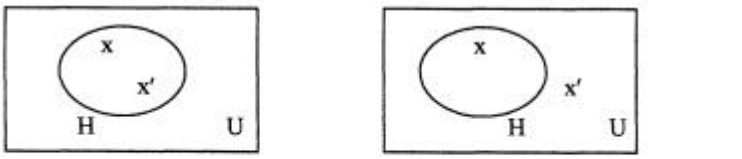
假設檢驗推理能力的心理學研究

「肯證偏誤」現象

心理學家發現人們在對所形成的假設進行檢驗時，普遍有一種「肯證偏誤」（confirmation bias）的傾向（e.g., Evans, 1989; Fischhoff & Beyth-Marom, 1983; Wason, 1960, 1966），也就是人們會尋找支持自己假設的證據，而較不會尋找有機會否定自己假設的證據。例如有研究發現人會重新解釋或棄之不顧那些與他們現存的假設不符的新訊息（e.g., Lord, Ross, & Lepper, 1979; Nisbett & Ross, 1980; Ross & Lepper, 1980）；或是人們會檢驗一些不太可能去否定他們假設的偏誤實驗（e.g., Snyder, 1980; Snyder & Campbell, 1980; Snyder & Swann, 1978）。這個現象首先在 Wason（1960）的研究中被提出。Wason 設計一個規則發現（rule discovery）的作業－「2 4 6 作業」，這是後來研究假設檢驗能力所經常使用的典範性作業。他要求受試者進行假設檢驗以發現一個實驗者預設的規則（例如，由小排到大的數列）。在作業的一開始，受試者只被告知符合這個規則的例子都由三個項目組成，而其中一個符合的例子是「2、4、6」，受試者可據此形成對預設規則的猜測（亦即形成假設）。隨後，受試者在接下來的測試中提出一組組的例子來測試其猜測。例如受試者提出「4、6、8」這組例子，來測試他的假設「偶數」。在每次測試之後，實驗者都會依照此例子是否符合預設規則而給予「是」或「否」的回饋。在受試者自覺很有信心後，可向實驗者說出他的猜測。Wason 在此作業中發現，79%的受試者在幾次測試後就很有信心地公佈其猜測，但是答案卻是不正確的，這些失敗的受試者傾向測試符合自己原有假設的例子或訊息，而較不會尋求可能不符合自己假設的例子。例如若受試者形成的假設是「偶數」，則傾向用偶數例子（如4、6、8）來測試，而較不會用非偶數例子來測試（如 1、3、6），前者通常被稱作「正例法」，後者稱為「反例法」[1]。圖 1 為這兩種測試策略的圖示。

「否證論」的觀點

上述的傾向之所以被稱為是一種「偏誤」，乃源自於 Popper（1959）的否證論。他認為得



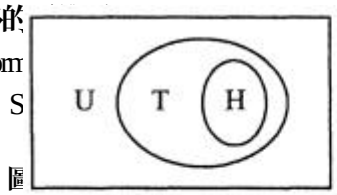
正例法：受試者根據起始例子 x 作出假設 H，然後選擇符合 H 的例子 x′來測試（U 為全部例子的集合）

反例法：受試者根據起始例子 x 作出假設 H，然後選擇不符合 H 的例子 x′來測試（U 為全部例子的集合）

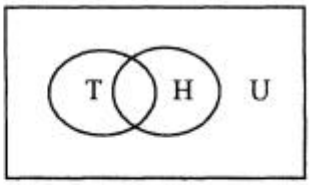
圖 1：正例法與反例法圖示

到支持的證據並無法證明假設或理論為真，只有不支持某假設的證據才能提供此假設確切的訊息（亦即此假設是錯誤的），因此得到否證才是一個科學理論或假設得以修正的要素。他的看法指出邏輯經驗論者的一個盲點，亦即以往經驗的歸納無法邏輯地得出未來的必然性，正如牛頓力學在過去千百次的檢證中都成立，但僅一次的日全蝕，就可顯示其不足。據此看法，一個好的檢測假設的策略應是盡量找出可以否證此假設的證據或例子。許多研究者為了提高假設檢驗作業的正確率，嘗試利用口頭指導的方法要求受試者特別注意在檢驗假設的過程中提出反例（e.g., Gorman, 1986;Gorman & Gorman, 1984; Gorman, Gorman,Latta & Cunningham, 1984; Gorman, Stafford& Gorman, 1987; Mynatt, Doherty & Tweney,1977; Tweney, Doherty, Womer, Pliske, Mynatt,Gross & Arkkelin, 1980），或是改變作業的程序，以直接或間接的方式增加否證的訊息來提高作業的成功率（Green, 1990; Klayman &Ha, 1989; Rossi, Cavemi & Girotto, 2001），但卻沒有一致性的效果。

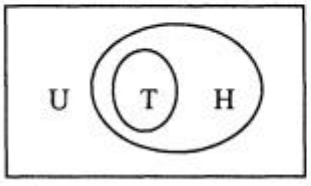
我們也發現過去研究者對於「否證」的定義並不一致，有些研究者認為多使用反例測試法，便可得到較多的 *et al*., 1986; Gorm Halberstadt & S



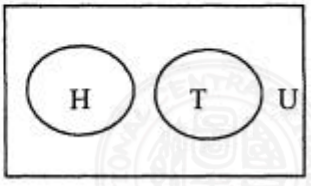
內含（embedded）



交疊（overlapping）



包圍（surrounding）



錯開（disjoint）

Mynatt *et al*., 1977; Penner & Klahr, 1996; Tweney *et al*., 1980）；另一些研究者（Gorman *et al*., 1984; Gorman *et al*., 1986; Gorman *et al*., 1987）將「否證」定義為「否」的回饋，即實驗者指出所測試之例子不符合預定之規則（disconfirm the correct rule）。除此之外，Klayman 和 Ha（1987）根據 Popper 的觀點認為「否證」即是所得到的回饋不符合受試者的預測，也就是當受試者測試一個正例，卻得到「否」的回饋，或是當測試一個反例，卻得到「符合」的回饋，如此的訊息才能否定受試者的原假設。他們進一步指出若把一個假設當成是其正例的集合，則受試者最開始形成的假設（初始假設）與預設或正確規則之間的關係，依據兩集合間交集的關係可分為四種（見圖 2），分別為內含（embedded，初始假設被包含在正確規則之內）、交疊（overlapping，初始假設和正確規則有部分交集）、包圍（surrounding，初始假設包含正確規則）與錯開（disjoint，初始假設和正確規則兩集合間沒有交集）。根據他們的分析，在某些合理的前提下，測試正例在機率上比反例法更易得到否證，因此不失為一種有效的測試法。例如在「2 4 6作業」中，受試者一開始形成的假設（H）為

由小排到大的數

此時若用正例

這個假

設，則實驗者會回答「否」，推理者即可得到否證原假設的訊息。但是在「2 4 6作業」中，過去的研究發現受試者根據初始例子「2、4、6」所想的初始假設常是「遞增偶數」，因而與預設規則是內含的關係，在這種情況下，正例法並無法有效地得到否證，因為任何「遞增偶數」的正例(如：「18、20、22」；「106、108、110」..)，也都同時為「由小排到大的數列」集合中的正例，因此不可能得到否證的訊息。

過去強調否證重要性的研究雖然很多，但並沒有任何研究分析過成功發現規則者與失敗者是否在 Popper 定義下的否證量有所差異。連韻文等人（連韻文, 1998; Lien & Lin, in preparation）首度針對 Popper 所提之否證進行分析，以受試者在整個測試過程中（共十二次測試）得到否證的次數為否證量，結果發現台大學生在「2 4 6作業」中，成功者與失敗者所獲得的否證量並沒有差異，顯示是否得到較多的否證訊息似乎不是區別成功者與失敗者的關鍵因素。

另有假設與假設檢驗推理表現

上述對於否證的討論主要著重於測試假設的過程中，如何能得到對於原假設有效的訊息（否證）。但除了測試假設的階段（testing phase），假設檢驗推理的歷程還包括產生新假設的階段（generation phase）。過去研究者也有不少人認為新假設的產生才是規則發現的重點。例如 Platt（1964）曾強調有效的科學研究需要評量多種可能的另有假設（alternative hypothesis）；Kuhn（1970）在其「科學革命的結構」一書中指出促成常態科學典範遷移（Paradigm Shift）的要件不在累積多量的異例(anomalies)，而是必須要有好的取代假設。Platt 與 Kuhn 的看法點出在測試過程中，除了得到否證以外，產生出可取代的另有假設（亦即新假設）是成功發現規則另一個重要的因素。得到否證訊息能夠淘汰掉錯誤假設，但若沒有適當的新假設取代，仍然不可能成功。於是有些研究者利用作業程序的改變以增加受試者產生另有假設的個數，以期能提高受試者在規則發現作業的表現。例如要求受試者每次測試兩個或四個假設(Laughlin & Bonner, 1999; Laughlin, Bonner & Altermatt, 1998; Tweney *et al.*, 1980)；或在測試前先寫下所有可能的假設(Green, 1990; Klahr & Dunbar, 1988)；或是分析作業中的成功者與失敗者在所產生另有假設的個數上有無不同（吳庭瑜、吳明樺和洪瑞雲, 1998; Klayman & Ha, 1989），但另有假設個數對假設檢驗推理的表現並沒有穩定的正面效果。

區分另有假設為交疊性另有假設與互包性另有假設

為什麼增加另有假設的個數也無法一致提高正確率呢？在過去一系列的研究中（連韻文, 1998; Lien & Lin, in preparation），我們發現推理者在假設檢驗推理過程中所產生的另有假設，可進一步區分為兩類。其中一類新假設是原有假設範圍的擴大或縮小（新舊假設間為互相包含的關係）；另一類則是形成新角度的假設（新舊假設的集合只有部份重疊），我們稱前者為「互包性另有假設」，後者為「交疊性另有假設」。上述的區分與 Klayman 等（1987）區分「預設規則集合」和「初始假設集合」間的關係類似（請參見圖 2），但在此所著重的是新產生的假設集合和前一個假設集合間的關係，並非和預設規則集合的關係。我們認為因為受試者並不知道預設規則為何，比較預設規則集合與假設集合並不能反映出受試者假設改變的歷程。例如，若受試者對第一個例子「2、4、6」所產生的假設是「偶數」，之後改為「整數」，則新假設集合包圍舊假設（為舊假設的範圍擴大，因為所有的「偶數」都屬於「整數」）；若受試者的新假設為「二位數以內的偶數」，新假設則內含於舊假設之內（為舊假設的範圍縮小，因為所有的「二位數以內的偶數」都屬於「偶數」），以上兩種皆為原集合範圍的擴大或縮小，我們合稱這

種新的另有假設（「整數」或「二位數以內的偶數」）為前一假設（「偶數」）的「互包性假設」。若受試者的新假設為「等差數列」，則新假設與舊假設之間的案例集合則只有部分重疊（交疊的部分如「2、4、6」，「30、26、22」.,. 而「3、6、9」只屬於「等差數列」，「12、90、36」只屬於「偶數」），此新假設（「等差數列」）即為前假設（「偶數」）的「交疊性假設」（圖 3 圖示這兩類假設與前一假設的關係）。

對於較複雜的現象或較困難的問題，我們常常不容易一開始就能想到逼近真象或答案的假設，因此這些問題才尚未解決（科學研究的主題常是屬於這一類的），也因而需要從新的角度來思考，若只產生互包性另有假設，只在原有概念範圍內搜尋，對幫助規則的發現效果有限。得到否證訊息卻失敗的受試者，可能就是陷入僅對原假設作範圍修正（也就是僅產生互包性的另有假設）的循環中。這樣的區分也呼應了科學哲學家 Lakatos（1970）繼 Popper 之後提出的「精緻否證論」。他認為科學的理論可分為兩個部份，第一為硬核，是整個研究綱領的基礎部份，具有不可動搖的特性；第二為保護帶，是一個巨大的輔助假說，頑強的保護硬核不遭到反駁。當遇到與理論相違的證據或現象時，科學家常抗拒修改理論的核心部份，而只在保護帶上變化。例如牛頓力學中的三大定律及萬有引力定律就是牛頓綱領的硬核部份，發現反常的現象後，只修改了保護帶的部份。我們認為對原假設作範圍的修正類似於只對科學理論的保護帶做修正。過去研究將所有的新假設都視為另有假設，是未注意到這兩種假設的性質不同，而將它們混為一談。

這種區分也得到實徵性資料的支持。連韻文等人（連韻文, 1998; Lien & Lin, in preparation）的研究結果顯示，在「2 4 6作業」中成功發現規則的大學生在交疊性另有假設的個數上顯著多於失敗者，但在互包性另有假設或假設總數

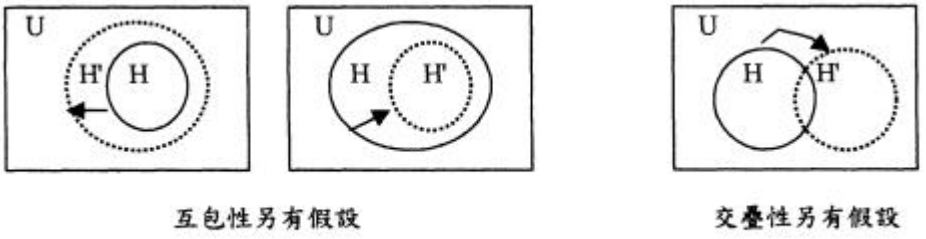


圖 3：交疊性另有假設與互包性另有假設之圖示

上並沒有差異；此外，在測試中得到否證訊息之後，成功者較失敗者顯著地產生較多交疊性假設，而失敗者則是在得到否證訊息後，較常對原假設作範圍擴大或縮小的修正，這樣的結果也對過去提高否證量與另有假設數皆無法有一致地促進效果提供了一個解釋。得到否證雖有助於刪除不正確的假設，也可促使推理者思索新假設，但若產生的新假設只是原有假設範圍的擴大或縮小，則無助於困難問題的解決。

測試策略

如上述，假設檢驗推理的歷程除了產生新假設外，還包含測試原假設這個部分。過去研究者在分析假設檢驗推理的測試歷程時，通常要求受試者寫出當下心中所認為的假設，以及接下來要繼續測試的例子，研究者在分析時即根據受試者欲測試的例子與該假設之間的關係（測試例子為假設的正例或反例）將測試策略分為正例法或反

例法(e.g., Gorman, 1986; Gorman *et al.*, 1984; Gorman et al., 1987; Mynatt *et al.*, 1977; Tweney *et al.*, 1980; Wason, 1960)。但Klayman 和Ha (1989) 在研究中要求受試者寫出他們測試每一個例子的理由。根據這些理由，他們發現有些受試者在測試時並不是單純的測試當下假設的正例或反例，而是又想出一個競爭假設，同時測試兩假設（例如，寫下「偶數」為可能規則，但測試後所寫的理由為想知道是「任意偶數」都可以還是一定要「遞增的偶數」？）。他們區分出此種測試方法為另有測試法（alternative testing strategy）。在我們過去一系列的研究中（連韻文，1998; Lien & Lin, in preparation），也發現受試者有時同時測試兩個假設的現象。不同於Klayman 等，我們將這種測試法稱為「雙假設測試法」(同時測試兩個假設)，以凸顯和正例法與反例法等單一假設測試法（只針對單一假設來測試）的不同。又因為雙假設測試法牽涉到兩個假設，如前述分析，兩假設的關係有必要再區分為交疊性關係與互包性關係，於是將「雙假設測試法」進一步再區分為兩類：互包性測試法（同時測試兩個有互相包含關係的假設）與交疊性測試法（同時測試兩個有部份疊合關係的假設)。各測試法的範例請參閱附錄一。使用雙假設測試法即是在尚未得到原假設的否證時，就主動想出不同的假設而同時測試。相較於得到否證之後，被迫想出新假設，多使用雙假設測試法中的交疊性測試法不失為一個主動提高交疊性另有假設個數的方法[2]。連韻文等人的研究發現成功者的確比失敗者使用更多的交疊性測試法，但在互包性測試法、正例與反例法的使用上並無差異。

綜合上述，在假設檢驗推理的測試階段，我們區分出兩大類四種策略：(1)單一假設測試法一僅測試某個假設是否正確，包括正例法或反例法。(2)雙假設測試法一同時測試兩個假設，包括交疊性測試法與互包性測試法。而在假設產生的階段方面，我們依照先後兩個假設的關係，將新假設區分為交疊性假設與互包性假設。圖4即顯示測試策略與假設種類的分類。



圖4：測試策略及假設種類之分類圖

研究目的

如前述，我們發現在假設檢驗推理的作業中，成功者與失敗者僅在交疊性另有假設與交疊性測試上有所差異，其他的測試策略（正例法、反例法、互包性測試法）以及互包性假設、假設總數和否證量上皆沒有差異，我們推測在較困難的題目上，是否能夠產生較多的交疊性另有假設可能是成功發現規則者與失敗者的主要差別。由於假設檢驗能力是科學知識學習與科學研究的內在歷程，一些受到社會認可具有科學研究潛力的人，他們在進行假設檢驗推理作業時是否採用較佳的策略或是具有上述成功者的特色？亦即是否能夠主動採用有效的交疊性測試策略、產生較多交疊性另有假設，而成功地找到正確的規則？在本研究的第一個實驗中，我們比較其科學研究潛力已受到認可的大學生（通過教育部甄選保送各大學基礎自然學科的科學資優生）與一般大學生（未參與甄選者）[3]在假設檢驗推理作業上的表現。我們預測這些具有科學潛能的人才，應更能夠成功地發現正確的規則，並且他們應該會主動採用較佳的測試策略（交疊性雙假設測試與產生較多的交疊性另有假設），以再一次檢證我們上述的看法，亦即影響假設檢驗推理表現的關鍵因素在於交疊性另有假設的產生，而非否證量的多寡。此外，經由這樣的研究，也可逐漸建立假設

檢驗或規則發現能力與科學學習成就的關係，以作為往後甄選類似人才的一項參考指標，另外也可進一步探究影響好策略與好假設產生的因素，而能將好的假設檢驗推理策略教導一般學生，以期能提升一般學生在假設檢驗推理方面的表現。因此在本研究的第二個實驗中，我們即要根據資優生的思考特質，探討提升一般生在假設檢驗推理表現的方法。

實驗一　科學資優生在假設檢驗推理上的表現

在本實驗中，我們以「2 4 6作業」分別測試一組通過「科學資優生甄試」的台大學生與一般通過聯考進入台大的學生，並比較這兩組大學生在成功率、形成的假設種類以及測試策略上的異同。我們欲探究這一群經過嚴格遴選，科學研究潛力受到肯定的大學生在假設形成與檢驗方面等能力上是否如我們所預期的，採用較佳的測試策略與形成較多的交疊性另有假設，因而有較高的成功率（找出預設的規則）。

方　　法

受試者

一共58位大學生參與本實驗，其中29位為84-87年度間經過教育部所舉辦「數學及自然科學資優生保送甄試」的考核（經過學校推薦、基本學科能力測驗、智力與性向測驗以及科學研習營的訓練與挑選）通過而進入台灣大學基礎科學的科系就讀的學生，包含了數學、物理、化學、地質、動物、植物、大氣等系。另外29名為台灣大學修習普通心理學的學生，應課程要求來參加實驗，共有文法或社會科學學院學生7人、管理學院學生3人、理工學院學生12人、醫學院學生4人以及農學院學生3人。

實驗材料與程序

每一組的受試者都單獨給予「2 4 6作業」。開始時，受試者被告知有一個預定的規則寫於一張紙上，裝於其面前的一個信封內，受試者的任務就是要去猜出此預設規則（此規則即為「由小排到大的數列」）。實驗者會先給予受試者一個符合此規則的正例：「2、4、6」，並告知此預設規則所衍生的例子都由三個項目所組成。受試者先根據此正例形成第一個假設，然後進行12次的測試，在每一次測試中，受試者先提出一個自認為對猜出正確規則有幫助的例子，然後寫下他為何測試此例子的理由，接著實驗者會根據所測試的例子是否符合預設的規則而給予「是」或「否」的回饋。例如受試者測試「8、16、20」，因為此例符合預設的規則（由小排到大的數列），實驗者回饋「是」；若受試者測試「8、20、16」，實驗者則回饋「否」。接著受試者寫下他目前認為最有可能的規則。我們所採用的作業程序和Klayman等人（1989）相似，而和Wason的原設計有以下二點不同。第一、每次測試受試者要另外寫出測試此例的理由，還要寫下目前他認為最可能的規則。這樣的程序可以方便實驗者分類出受試者所使用的測試策略；第二、完成作業的時間沒有限制，但不管猜出的規則對否，受試者要做完規定的12個測試後才結束作業。Wason的原作業程序不限定測試次數，如此推理者容易受到動機等因素影響，而使測試不足、訊息不夠，較不容易觀察到較全面的策略使用，也使分析者無法有足夠的訊息分析。

除了告知所測的例子是否符合規則外，實驗者未給予任何其他回饋，在過程中也不告知受試者所猜的假設是否正確，直到12次測試皆結束後才公布最後正確的規則。通常完成作業所需時間約30分鐘。

在正式作業之前有一個練習題目，所給予的正例為「印尼、菲律賓、馬來西亞」，受試者據此做兩次練習測試。練習題中沒有預設正確的規則，實驗者隨機給予「是」或「否」的回饋。

結　果

假設種類與測試策略的分類

首先我們根據前述的分析，統計每個受試者在十二次的測試中使用何種測試策略與所產生的假設種類。分類的方法如下：

假設種類　除了受試者根據所給的例子「2、4、6」所產生的第一個假設外，我們將受試者在後來的12次測試中所推測的假設與其前一個假設的關係區分為兩種–互包性假設與交疊性假設（判準見前述有關另有假設的分析）[4]。受試者所產生的假設總數即為此兩種假設之和加上初始假設。

測試策略　如前述，我們根據受試者在每次測試中所測試的例子與其測試理由，將其測試策略區分為四種（詳見前面有關測試策略的分析與附錄一）。若受試者以正例測試其目前的假設，則為正例測試法；若以反例測試心中假設，則為反例測試法。若受試者的測試例子想要區分兩個假設，則屬雙假設測試法，又可根據兩假設之間的關係分類為交疊性測試或互包性測試。有些測試原則上雖屬雙假設測試法，但無法提供有效的區分（例如想區分「偶數」和「等差數列」兩個假設，卻選擇「4、6、8」來測試，此例同時屬於「偶數」和「等差數列」），我們稱此為非關鍵性測試（此即為 Klayman 等人（1989）所稱之 failed alternative test）。根據我們的大學生資料，這類測試的次數都很少（12 次測試中平均次數少於 0.3 次，< 2.5%），不論是成功者與失敗者之間或是資優生與一般生之間，這類測試都沒有差異，因而在此不予報告。

以上分類的評定皆由兩評分者獨立進行，評分者間分類相同的比率為96.5%[5]，不同的部分最後亦經討論而得到共識。

成功率

如我們所預測的，資優生組的成功率（55.2%）顯著的高於一般生組[6]（20.7%），$\chi^2 = 7.32, p < 0.01$。

假設總數與種類

過去的研究認為提高否證量或增加假設總數和成功率的提升有關，我們則認為只有產生交疊性的新假設才是關鍵。實驗的結果大致支持我們的看法，但科學資優生的表現仍然有一些特色是過去我們在一般大學生成功者上沒有發現的。

在假設總數方面，科學資優生較一般大學生產生更多的假設（平均個數分別為 10.21 與 8.03, *SDs* = 1.90 與 3.45, $t_{43.5} = -2.97$[7], $p = 0.005$, $d = 0.78$[8]）。但如我們所預期的，若進一步區分出交疊性假設與互包性假設，則資優生在交疊性假設較一般生多（平均個數分別為 1.55 與 0.52, *SDs* = 1.12 與 0.63, $t_{44.2} = -4.33$, $p = 0.00$, $d = 1.14$），但在互包性假設上與一般大學生沒有差異（平均個數分別為 7.66 與 6.52, *SDs* = 2.08 與 3.45, $t_{45.9} = -1.52$, $p = 0.14$）。表 1 呈現科學資優生與一般大學生所產生不同假設種類的平均個數和標準差。

表1：實驗一中，資優生組與一般生組所產生的兩種新假設之平均個數與標準差

假設種類	平均個數	
	資優生 (n = 29)	一般生 (n = 29)
交疊性另有假設	1.55 (1.12)	0.52 (0.63) **
互包性另有假設	7.66 (2.08)	6.52 (3.45)

註：括弧中的數字為標準差。** p < .01。

表2：實驗一中，資優生組與一般生組使用四種測試策略的平均次數與標準差

測試法		平均次數（次/12 嘗試）	
		資優生	一般生
單假設	正例測試法	2.59 (2.08)	5.00 (3.68)**
	反例測試法	1.66 (1.63)	1.48 (1.62)
雙假設	交疊性測試法	0.72 (0.75)	0.31 (0.71)*
	互包性測試法	6.69 (2.65)	4.97 (3.46)*

註：資優生組與標準組各有29人。括弧中的數字為標準差。
* p < .05，** p < .01。

測試策略

在測試策略方面，過去的研究者認為多使用反例法有助於正確率的提升，但本實驗結果和我們過去的研究同樣地不支持此種看法。資優生與一般生使用反例法的平均次數沒有差異（平均次數分別為 1.66 與 1.48, *SDs* = 1.63 與 1.62, t_{56} = −0.40, p = 0.69）。一個不在預期之中但不令人意外的發現是資優生顯著地較一般生常使用雙假設測試法（平均次數分別為 7.41 與 5.28, *SDs* = 2.87 與 3.60, $t_{53.3}$ = −2.50, p = 0.016, d = 0.66）。進一步分析發現相對於其他大學生，資優生使用較多的交疊性測試法（平均次數分別為 0.72 與 0.31, *SDs* = 0.75 與 0.71, t_{56} = −2.15, p = 0.036, d = 0.57）與互包性測試法（平均次數分別為 6.69 與 4.97, *SDs* = 2.65 與 3.46, $t_{52.4}$ = −2.13, p = 0.038, d = 0.56）。而在正例法方面，資優生則顯著地少於一般生（平均次數分別為 2.59 與 5.00, *SDs* = 2.08 與 3.68, $t_{44.2}$ = −3.07, p = 0.004, d = 0.81）。表 2 呈現資優生與一般生在 12 次測試中使用不同測試策略的平均次數和標準差。

否證量

如前述，過去研究者對於否證的定義並不一致，若根據 Klayman 等（1987）的分析來計算否證量（即當受試者測試一個正例，卻得到「否」的回饋，或是測試一個反例，卻得到「符合」的回饋），資優生與一般大學生所得的否證量並無差異（平均次數分別為 1.38 與 1.38, *SDs* = 1.32 與 1.66, t = 0）。但在本實驗中，資優生顯著地傾向使用雙假設測試，我們認為受試者使用雙假設測試時，若選擇一個關鍵性的例子區分這兩個假設，可以得到其中一個假設的否證和另一個假設的肯證，因此應該將關鍵性的雙假設測試合併入否證量的計算比較合理。根據我們過去的研究結果，成功者與失敗者不論在僅視單假設測試或是加入雙假設測試計算的否證量上皆沒有差異，但在此若合併雙假設測試所得到的否證量計算，資優生顯著得到較一般生多的否證量（平均次數分別為 8.79 與 6.66, *SDs* = 2.16 與 3.30, $t_{48.3}$ = −2.92, p = 0.005, d = 0.77）。

討　論

資優生推理的特色

雖然資優生在假設檢驗推理的作業上確實較一般大學生有好的表現，也如我們之前所分析與預測的較一般生常採用交疊性測試法，並較一般生產生較多的交疊性另有假設（亦即具有典型成功者的特色），但資優生與一般生之間也有一些差異是我們過去比較一般生成功與失敗者之間所沒有發現的。例如，科學資優生較少使用正例法、較常採用雙假設測試法，並且產生較多的假設總數。另外，若合併雙假設測試所能得到的否證量，資優生也較一般生得到較多的否證。這些差異和成功發現規則的關係如何呢？從資料看來，假設總數的差異，可明顯看出是因為資優生較一般生產生更多的交疊性假設，而互包性另有假設個數沒有顯著差異（假設總數＝交疊性另有假設個數＋互包性另有假設個數＋1）。而資優生較少採用正例法可明顯看出是因為他們較傾向使用雙假設測試策略（等於交疊性測試次數＋互包性測試次數）的緣故。由於實驗中每位受試者的測試次數限制為 12 次，雙假設測試法的增加必然會減少其他兩種測試法（正例與反例法）的使用次數，由於一般生正例法的使用遠比反例法為多，而使得相較之下，資優生正例法次數下降的幅度就較為顯著。另外，我們認為資優生得到較多的否證量可能也是因為他們傾向採用較多的雙假設測試，因此，比較值得注意的反而是雙假設測試法的使用（將在下面段落討論）。此外，我們也計算了各個顯著差異的效果量 d，結果顯示交疊性假設的效果量非常大（d = 1.14），也較其他差異為大（參見前列結果）。

我們也進一步比較所有受試者（合併資優生與一般生）中成功歸納出規則與失敗者的差異，成功者也顯著地較失敗者產生較多的交疊性另有

假設（平均值分別等於 1.45 與 0.78, *SDs* = 1.18 與 0.87, $t = 2.51$, $p = 0.015$, $d = 0.68$），但在互包性假設與假設總數上並無顯著差異（互包性假設平均值分別等於 7.22 和 7.00, *SDs* = 2.20 與 3.25, $t = 0.29$, $p = 0.78$；假設總數平均值分別等於 9.68 和 8.78, *SDs* = 2.06 與 3.39, $t = 1.26$, $p = 0.21$）。否證量方面也同樣地無顯著差異（僅算單假設測試：平均值分別等於 1.00 和 1.61, *SDs* = 1.38 與 1.52, $t = 1.54$, $p = 0.13$；合併單、雙假設測試：平均值分別等於 8.32 和 7.36, *SDs* = 1.91 與 3.43, $t = 1.36$, $p = 0.18$）。除了失敗者正例法的使用顯著地多於成功者（平均值分別等於 4.50 和 2.64, *SDs* = 3.61 與 1.99, $t = -2.53$, $p = 0.014$, $d = 0.69$）外，其餘測試法兩組並無差異。這些結果和我們過去的研究結果吻合，再度支持我們的看法，顯示只有交疊性另有假設是在「2 4 6作業」上表現好與不好的穩定區別，其他的差異有可能只出現在某些成功者身上。

雙假設測試法的可能優勢

資優生很明顯地從一開始就較一般生更常採用雙假設測試。理論上，使用雙假設測試有可能只增加互包性假設，而非交疊性假設，而交疊性假設的產生也不一定要透過雙假設策略的使用（許多一般生的成功者並不常使用雙假設測試），但我們認為採用雙假設測試策略可能可以有效地促進交疊性另有假設的產生。細看資優生所做的雙假設測試，不但如我們所預期地較一般生更傾向採用交疊性測試，也超乎預期地較常採用互包性測試。但如果進一步分析交疊性測試與互包性測試使用的時機，則兩者有很大的不同。我們將 12 次測試分成前、中、後三個區段（分別為第 1-4 次測試、第 5-8 次測試、第 9-12 次測試），分別計算資優生與一般生在每個區段所採用的各種雙假設測試次數。結果發現（如圖 5 所示）在交疊性測試方面，資優生組在三個區段的分佈有顯著差異（平均次數分別為 0.69、0.14 與 0.07, *SDs* = 0.71、0.35 與 0.26, $F_{1.4,38.2} = 15.72$, $p < 0.01$）。LSD 事後比較顯示前區段的次數顯著地較後兩個區段多（p 值皆小於 0.01）。而一般生組在交疊性測試的使用方面，三個區段都相對的少，且其間沒有顯著差異（平均次數分別為 0.17、0.10 與 0.03, *SDs* = 0.47、0.41 與 0.19），因而資優生與一般生在交疊性測試的分佈上有交互作用（$F_{1.3,74.9} = 6.64$, $p < 0.01$），亦即資優生在交疊性測試的次數在前區段顯著高於一般生組，但在後兩個區段，兩組受試者則無顯著差異。

至於互包性測試的分佈則全然不同。在資優生組方面，三個區段間也有差異（平均次數分別為 1.83、2.59 與 2.38, *SDs* = 1.28、1.21 與 1.29, $F_{2,56} = 3.55$, $p < 0.05$），基本上是較後區段的次數高於前區段。LSD 事後比較顯示前區段的互包性測試次數顯著少於中區段（$p < 0.05$），其餘比較則無顯著差異。一般生組也有同樣的傾向（平均次數分別為 1.31、1.83 與 2.03, *SDs* = 1.23、1.61 與 1.48, $F_{2,56} = 3.86$, $p < 0.05$），LSD 事後比較顯示前區段的互包性測試次數顯著少於後區段（$p < 0.05$），因此資優生組與一般生組在互包性測試的分佈上並無交互作用。

這樣的結果顯示了在測試初期若採用雙假設測試可能較易提高交疊性假設的產生，若在測試的後期使用雙假設測試，相對上更容易產生兩個具互包性關係的假設。資優生由於在整個測試過程中持續使用雙假設測試法，因而交疊性測試與互包性測試都較一般生為多。相反地，一般生則很少使用雙假設測試，尤其是在前區段。這個新發現並不令人意外，過去心理學家發現人在問題解決的歷程中有所謂的「心向效應」（set effect）和功能固著（fixation）的現象，即對一個問題經由重複的解題經驗產生一個解題的心向（mental set）或對某物的功能產生一固定的認識之後，就很難用另一種可能是更簡單的策略來解類似的問題或是很難想到某物有另外的用法，這意味著人若反覆在某種思考之中，就極容易陷入相同模式而無法跳脫。同理，在假設檢驗推理的過程中（也是一種問題解決），若推理者心中已經形成一個可能的候選假設，隨著肯定此一假設的次數增加，推

理者在遇到否證後，思索取代假設時，就可能越不易跳脫原假設的範疇。因此在測試初期採用雙假設測試的可能好處是及早促進交疊性假設的產生，無須等到後來否證的出現時，已不易想到新角度的假設了。

在測試後期使用較多的互包性策略（使用

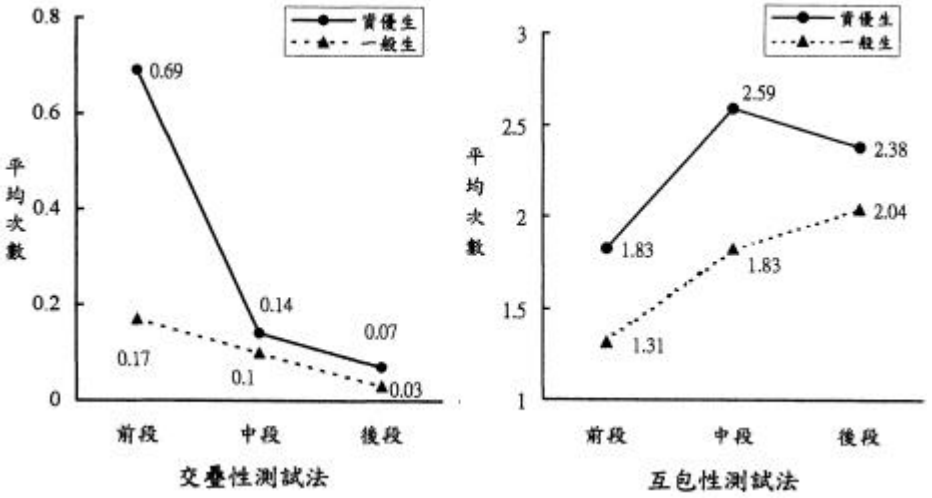


圖 5：資優生與一般生在前、中、後三個區段中使用雙假設測試策略之圖示

雙假設測試法時）或產生較多的互包性假設其實也是必要的。基本上，當假設已經很接近答案時，就需要這種局部微調的過程。不過這種能力似乎並不是成功者所獨有，從本實驗或我們過去研究的資料看來，並未發現互包性測試或互包性假設可以穩定地區分成功者與失敗者。但不可否認的，本實驗並無法排除資優生的優異表現是互包性測試與交疊性測試同時增加的共同效果。但若我們上述的分析是對的（在測試前期若使用雙假設測試法較容易產生與原假設是交疊性關係的新假設），則教導受試者在前區段使用雙假設測試策略應該可以單獨提高交疊性測試次數與交疊性假設的個數，且在互包性測試沒有顯著變化的情況下提高正確率。若此，則可進一步支持我們的論點。我們將在下一個實驗中驗證此論點。

實驗二　提升假設檢驗推理表現的指導法—「有限的雙假設測試法」

從資優生的實驗結果中可以看出，使用雙假設測試策略應該有它的正面效果，但過去的研究要求受試者每次測試兩個或四個假設卻曾報告相反的結果（Laughlin & Bonner, 1999; Laughlin, Bonner & Altermatt, 1998; Tweney *et al.*, 1980）。Tweney *et al.* (1980)解釋這樣的策略對受試者而言太過困難，受試者較能掌握單假設的測試。但是我們認為這個困難可能是來自於全程使用雙假設測試對一般大學生而言工作記憶（Baddeley, 1976）負擔太大。使用雙假設測試策略必然會較使用單假設測試策略耗費更多的工作記憶資源，因為推理者必須形成另一個假設、同時記住兩個假設、再選擇一個具區辨性的例子進行測試，尤其當所形成的兩個假設是交疊性的關係時會較形成互包性關係的假設更為困難（連韻文, 1998）。資優生會頻繁地採用這種有效但須較多工作記憶的雙假設測試策略，可能與他們具有較大的工作記憶廣度有關，故較一般學生更能輕鬆、主動地使用。

資優生的工作記憶廣度

我們邀請本實驗其中12位資優生[9]再回來進行「工作記憶廣度測驗」，確實發現資優生在工作記憶廣度的測驗分數顯著高於一般生。我們所採用的工作記憶廣度測驗包括中文閱讀廣度測驗（陳皓照, 1998）與數字工作記憶測驗[10]兩個部分，都以電腦呈現。每位受測者都先後接受兩個測驗，測驗順序隨機決定。在中文閱讀廣度測驗中，受試者要一邊朗讀一系列中文句子，一邊在心中記下每一句子最後所標明要記憶的雙字詞，朗讀完指定句數的句子後，必須依序回憶出目標雙字詞，指定朗讀的句數由少至多，分別為2、3、4、5、6句。數字工作記憶測驗是在電腦中依次呈現三個代數式的畫面，分別為A與B的關係、B與C的關係、C的數值，每個畫面持續2秒鐘後消失，接著要求受試者按照順序回答出B、A、C的正確數值（例題見附錄二）。此兩分測驗設計的原理都是要求受試者同時進行記憶和心智運算的雙重工作，若工作記憶的廣度越大，作業的表現應會越好。結果顯示科學資優生在語文廣度測驗的平均得分顯著高於105位台大的一般生（平均分別為4.08和3.31, *SDs* = 1.16和0.88, $t_{115} = -2.79$, $p < 0.01$, $d = 0.85$）。在數字測驗方面，平均得分也顯著高於一般生（平均分別為64和56.26, *SDs* = 7.27與9.37, $t_{115} = -2.77$, $p < 0.01$, $d = 0.84$），符合我們的預期。另外我們也發現此12位資優生在「246作業」的成敗與工作記憶測驗分數的高低有顯著的相關（$r = 0.71$, $p < 0.01$），這樣的結果再次支持我們的想法，假設檢驗推理作業的成敗與工作記憶廣度有關。

交疊性另有假設的產生和工作記憶廣度的關係

過去的心理學研究也在其他牽涉到不同認知能力的作業中發現工作記憶廣度大，可同時考慮多個不同向度的線索及假設，而有助於各種推理。例如Johnson-Laird（1983）指出受試者在做較困難的三段式演繹推理題目時（佔用較多工作記憶的資源），不易於同時形成兩個心理模型（mental model）來表徵關係。在語言理解方面，研究者也發現工作記憶廣度大者在理解模糊語句的過程中較廣度小者更能同時保有一個以上的解釋（Just & Carpenter, 1992）。我們（連韻文, 1998; Lien & Lin, in preparation）則認為工作記憶負擔的增減對交疊性另有假設的影響比對互包性假設為大。我們曾將同一母群的大學生隨機分派到三組工作記憶負擔不同的組別，分別進行「246作業」。其中標準組所做的「246作業」如同上述；加重負荷組的作業格式同於標準組，但同時必須隨著節拍由一千向上數數；而減輕組則是將其受試者的記錄全部放在同一頁中，而非隨著不同測試而記錄在不同頁中，因而減輕推理時的記憶負擔。結果顯示，工作記憶減輕組的正確率（45%）顯著的高於加重負荷組（10%），而標準組（20%）則介於兩者之間。更重要的是減輕組所產生的交疊性假設個數顯著高於加重負荷組與標準組，而互包性假設則三組之間沒有差異。在測試策略方面，三組並無不同。此結果顯示交疊性另有假設的產生受工作記憶負擔的影響較大。因此若一般大學生被迫全程使用雙假設測試法，可能會因認知負擔過重，反而增加大量的互包性假設或錯誤的假設而亂了陣腳。

有限的雙假設測試法

整合上述幾個發現，我們推測若只要求一般生在前幾次的測試中利用雙假設測試法進行測試，則在要求的假設個數上較不會超過受試者的認知負擔，在時機上也是受試者想法較具彈性的時候，應該可以顯現出類似資優生使用雙假設測試策略的優點，提高受試者產生交疊性另有假設的機率，改善受試者在假設檢驗方面的表現。根據上述想法，在本實驗中我們設計「有限的雙假設測試法」。我們指導一組大學生使用「有限的雙假設測試法」來進行「246作業」。相較於另一組來自同樣母群而未接受此指導的大學生，我們預測指導組會產生較多的交疊性另有假設，而有較高的成功率。

方　法

受試者

受試者為台灣大學修習普通心理學的學生，應課程要求來參加實驗，隨機分派至指導組（22人）或對照組（19人）。採用受試者間設計。

實驗材料與程序

對照組與指導組的作業與程序同實驗一，受試者根據「2、4、6」此一正例形成初始假設，然後同樣地有12次測試機會。在每一個測試例子之後同樣要在記錄表格上寫出測試此例的理由、記錄實驗者的回饋（是或否）、然後寫下目前他認為最可能的規則。但指導組在給予正例「2、4、6」之後，指導語中額外要求受試者「請你根據這個例子推想出二個有關此規則的假設。… 得到回饋後，你可以維持或修改你的假設，但你在前四個測試中，仍然必須維持能提出兩個可能的假設。」

結　果

成功率

如我們所預測，指導組的成功率（81.8%）極顯著的高於對照組（26.7%），$\chi^2 = 15.14$，$p < 0.01$。

假設總數與種類

假設種類與測試策略的分類同實驗一。如我們所預期，指導組的假設總數因指導語所致而顯著多於對照組（平均個數分別為8.73與6.21, *SDs* = 2.71與3.39, $t_{39} = -2.64$, $p < 0.05$, $d = 0.83$）。進一步分析發現指導組所產生的交疊性另有假設遠多於對照組（平均個數分別為2.45與0.68, *SDs* = 1.68與0.89, $t_{32.7} = -4.30$, $p < 0.01$, $d = 1.35$），但在互包性另有假設上，兩組並沒有差異（平均個數分別為4.91與4.53, *SDs* = 3.1與3.55, $t_{39} = -0.37$, $p = 0.71$）。表3呈現指導組與對照組所產生不同假設種類的平均個數和標準差。

表3：實驗二中，指導組與對照組所產生的兩種新假設之平均個數與標準差

假設種類	平均個數	
	指導組 (n = 22)	對照組 (n = 19)
交疊性另有假設	2.45 (1.68)	0.68 (0.89) **
互包性另有假設	4.91 (3.10)	4.53 (3.55)

註：括弧中的數字為標準差。** $p < .01$。

在否證量方面，指導組與對照組間沒有顯著的差異（僅計算單假設測試後得到的否證數，指導組與對照組平均值分別等於0.55和1.10, *SDs* = 0.91和1.15, $t_{39} = 1.74$, $p = 0.09$；合併單、雙假設測試後，平均值分別等於6.68和5.11, *SDs* = 2.57與3.09, $t_{39} = -1.78$, $p = 0.08$）。

測試策略

在受試者所使用的測試策略方面，由於指導組必須使用至少四次的雙假設測試法，我們預測指導組將較對照組採用較多的雙假設測試法，結果如我們所預期（平均次數分別為6.14與4.00, *SDs* = 2.73與3.07, $t_{39} = -2.36$, $p < 0.05$, $d = 0.74$）。

此外，如同我們之前的分析，當在前區段的測試中就要求推理者使用雙假設測試策略時，推理者所形成的雙假設較有機會是交疊性的關係，而非互包性的關係，換句話說，較有可能使用交疊性測試策略，而非互包性測試策略。結果符合我們的預測，指導組在交疊性測試的使用上顯著多於對照組（平均次數分別為2.36與0.37, *SDs* = 1.94和0.6, $t_{25.5} = -4.58$, $p < 0.01$, $d = 1.43$），但在互包性測試上並無差異

表4：指導組與對照組使用四種測試策略的平均次數與標準差

	測試法	平均次數（次/12嘗試）	
		指導組	對照組
單假設	正例測試法	2.41 (1.87)	5.84 (3.52)**

	反例測試法	1.64 (1.36)	1.95 (1.81)
雙假設	交疊性測試法	2.36 (1.94)	0.37 (0.60)**
	互包性測試法	3.77 (2.98)	3.63 (3.24)

註：指導組與對照組分別為22人與19人。括弧中的數字為標準差。** $p < .01$。

（平均次數分別為3.77與3.63, *SDs* = 2.98與3.24, $t_{39} = -0.15$, $p = 0.89$）。至於在單假設測試策略方面，對照組在正例法的使用上顯著多於指導組（平均次數分別為5.84與2.41, *SDs* = 3.52與1.87, $t_{26.5} = 3.82$, $p < 0.01$, $d = 1.20$），反例法的使用兩組並無差異。此外，指導組的受試者也較對照組的受試者發生更多次的非關鍵測試（平均次數分別為1.82與0.21, *SDs* = 1.53與0.42, $t_{24.56} = -4.72$, $p < 0.01$, $d = 1.48$）。表4呈現指導組與對照組在12次測試中使用不同測試策略的平均次數和標準差。

討　論

指導組的有限雙假設測試設計使得受試者的成功率大幅提高到80%以上，而較對照組高了60個百分點，且指導組的受試者確實較一般生採用更多次的交疊性測試法，產生了較多的交疊性另有假設，這與實驗一中的資優生有相同的特色。與實驗一中的資優生表現不同的是，指導組的受試者並沒有使用較多的互包性測試策略。誠如我們先前的分析，資優生可能是因為具有較大的工作記憶廣度，所以頻繁地採用較花費工作記憶資源的雙假設測試策略，他們在測試的前期傾向產生交疊性的新假設進行測試，後期則產生互包性的新假設。而一般學生可能因為工作記憶的限制較不會主動採用雙假設測試法，但若要求他們在前四次測試都採用雙假設測試策略，他們則會形成較多的交疊性假設，是以指導組中的受試者與資優生相同會有較多的交疊性測試法，而不若資優生採用更多互包性測試法。我們進一步計算指導組前四次雙假設測試中交疊性測試法所佔的比率，發現交疊性測試法所佔的比率為58.3%，若將整個測試過程納入計算，則交疊性測試佔了全部雙假設測試中的38.5%，而未經指導的對照組，交疊性測試法僅佔全部雙假設測試中的9.25%。這樣的結果再次支持了我們的預期，亦即在測試初期使用雙假設測試策略，可以提升交疊性測試的使用，顯示「有限的雙假設測試法」是一個有效的指導策略。

另外在否證量方面，即使是合併雙假設測試所得的否證量，指導組所得的否證量仍不多於對照組。為何和實驗一中資優生的結果有所差異？我們認為有可能是因為指導組的受試者只被要求前四次一定要做雙假設測試，不若資優生從頭到尾都傾向做雙假設測試，因此否證量不若資優生增加的多，但指導組的成功率卻大大增加（甚至高於資優生的成功率），可見否證量並不是成功或失敗的關鍵。這樣的結果呼應了我們過去的研究結果。

非關鍵性測試

結果也顯示了指導組較對照組發生更多次的非關鍵測試，亦即他們在測試前產生了兩個可能的假設，但是在測試此兩假設時，所選擇的測試例子不是同屬於兩個假設，就是同樣不屬於此兩假設，也就是測試例子無法區分兩個可能的假設究竟何者該捨棄、何者要留下。我們認為若受試者是主動採用雙假設測試策略（例如實驗一的資優生），他們在當次的測試目的就是要區辨出所想到的兩個可能假設，但在實驗二的指導組中，受試者是被要求一定要做出兩個假設，可能他們心中其實想先測其中一個假設，所以才會有時不經意選擇一個不具區辨性的例子（平均佔雙假設測試策略的24.58%）。不過根據我們的觀察，發生此情況的受試者多半在當次測試結束後，要下結論時就會發現這樣的問題，而在下一個測試就會選擇一個區辨性的例子來區別兩個可能的假設，所以在此情形下，某程度地增加非關鍵性測試次數似乎不影響正確率的提升。

得到正確答案的速率

主動採用交疊性測試除了能增加交疊性另有假設的產生外，我們發現這樣的策略也能夠讓受試者更有效率地發現正確的規則。我們分析經指導與未經指導[11]而成功的受試者答出正確規則的快慢，結果發現指導組的成功者快於未經指導的成功者（分別平均在第5.22次與第7.40次嘗試發現正確的規則，$p = 0.08$，達邊際顯著水準）。我們認為這可能是由於未經指導的對照組因為並未被要求使用雙假設測試法，他們修改假設的時機可能是在使用單假設測試並得到否證之後，而指導組的受試者因為在前四次測試中都必須想出兩個可能的假設，因而能較快在假設空間（hypothesis space）中搜尋到正確的假設。這樣的結果呼應了我們先前的分析，相較於得到否證之後被迫想出新假設，多使用交疊性測試法不失為一個主動且有效率提高交疊性另有假設個數的方法，也因而可能增快發現正確規則的速率。

「有限的雙假設測試法」的有效性

過去的研究者曾試過種種的方式以提高受試者在假設檢驗推理作業的表現，不是徒勞無功，就是必須大幅改變作業形態，例如觀看別人測試一個題目後再自行進行另一個假設檢驗推理的題目（Kareev et al., 1993）；或者改變作業目標，由歸納出一個規則改變成歸納出兩個互補的規則（例如遞增數列與非遞增數列）（Tweney *et al.*, 1980; Wharton, Cheng, & Wickens, 1993）。與其他方法相較之下，有限的雙假設策略是一個更為簡單而有效的方法，不但能大幅提高推理的正確率，更可能加快發現正確規則的速率。就教學上的意涵來說，在不超過學生認知上的負擔下，提醒學生先想出多種可能假設再進行測試，以逐一淘汰不正確的想法，比只提供單一假設的否証更能避免學生固著於錯誤的假設，更有機會建立正確的知識架構。

結　論

本研究根基於過去認知心理學有關假設檢驗推理的研究，利用假設檢驗推理的作業來探究：第一、已受認定具有科學研究潛能的人才在尋找規則的過程中有哪些特色，和一般大學生有何不同？他們使用哪些有效的策略？是否和我們過去發現的成功者有相仿之處？第二、是否能根據這些研究結果設計出有效的方法來幫助一般學生成功地找到規則？結果顯示，實驗一中經過教育部所舉辦「數學及自然科學資優生保送甄試」考核通過的資優生確實比一般大學生在假設檢驗推理的作業上有較高的成功率。他們也如所預測地較一般大學生產生更多的交疊性另有假設，具有我們過去研究中成功發現規則者的特色。除此之外，資優生在假設檢驗的過程中較一般學生更傾向使用雙假設測試策略，這樣的傾向有助於資優生在解決問題的初期產生較多的交疊性的假設。根據我們過去的研究，我們認為資優生能主動且頻繁地採用比單假設測試策略需要更多工作記憶資源的雙假設測試策略，可能和他們具有較豐富的工作記憶資源有關。我們的確也發現資優生在「工作記憶廣度測驗」上的表現顯著優於一般大學生。

在實驗二中，我們根據資優生的特色並考量一般學生工作記憶負荷量的限制，設計「有限的雙假設測試法」來指導一般學生，結果大幅提升一般大學生在假設檢驗推理作業的表現，顯示一般大學生只要在認知負荷不超過限制的狀況下，可以經由簡單的指導，而使用較佳測試策略，產生較多交疊性另有假設，並較快地找出預設規則。

上述的結果與我們過去的研究結果吻合，亦即成功發現規則者與失敗者關鍵的區別在於是否能夠產生新角度的交疊性另有假設，得到否證或互包性另有假設的多寡並不是兩者主要的分野。當然這並不是意涵否證對於規則的發現全然無足輕重，或是互包性另有假設就是沒有用的另有假

設。事實上，不論是肯證或否證都是在假設檢驗推理過程中必要的訊息，如果沒有肯證，推理者無法有任何依據形成可能的假設，更遑論接續的測試假設；否證則可以指出原假設的不足，提供修正假設的契機，也可淘汰掉不正確的假設；而當假設已經與目標範疇十分逼近時，推理者僅需在原假設的範圍上稍作修改就可找出正確的規則（即產生新的互包性的假設）。但我們認為當遇到較困難的推理問題時，推理者不易在一開始就想到逼近真相的解答，是否產生交疊性另有假設會是成敗的關鍵。

由於假設檢驗推理的能力與發掘科學真相、創造性的思考息息相關，在此一方面能力的評估，可成為往後甄選科學潛力人才的另一項參考。實驗一的結果也肯定了前述的甄選程序是可以有效挑出具有良好科學潛能的學生。

心理學研究在科學教育上的應用

將另有假設進一步區分成交疊性與互包性另有假設與教育界在探討「概念改變」時著重於「弱重建」與「根本重建」的區別（Carey, 1985, 1986; Vosniadou & Brewer, 1987）有相通之處。所謂「弱重建」與「根本重建」即相當於 Piaget 對知識建構機制所提出的「同化」（assimilation）與「調適」（accommodation）過程。「同化」為在原本的知識體系架構中添加入新習得的訊息，而「調適」則發生在新訊息無法同化入原本架構中而將原本的知識體系作一新的改變；因此，弱重建意謂著新事實的累積及現存概念之間新關係的形成，而根本重建意謂著核心概念的改變、結構的改變、及對現象之解釋的改變（見洪振方, 1996）。對於科學知識的學習，時常需要學生對原本的迷思概念作一根本的重建，而不僅是基於原本知識架構對新訊息做出理解，這也是為什麼科學知識的學習常讓學生感覺到困難。由這個角度看來，弱重建即相當於形成互包性的另有假設，僅將原本的假設因應新訊息的需要稍做範圍性的修改，而根本重建則相當於形成交疊性的另有假設，必須跳脫原假設的框架，以新角度解釋現存的所有證據。如我們之前所分析的，對於較複雜的現象或較困難的問題常需涉及交疊性另有假設的產生才較可能找出正確的解答。因此進一步探究影響兩種假設產生的因素，不但有助於學習者採用較佳的策略，也有助於教育者思考如何促使概念的根本重建。例如在促進概念改變教學法中所使用的「異例法」（discrepant event or anomaly）（Nussbaum & Novak, 1982）中包含了三個階段：第一階段、製造曝露事件（creation of an exposing event），引出主題要求學生表明立場；第二階段、引入異例，使學生明瞭自己的迷思概念與事實不符（亦即獲得否證），產生認知衝突；第三階段、調適期，教師鼓勵學生尋找答案來解決衝突。根據我們過去對於假設檢驗歷程的研究顯示（Lien & Lin, in preparation），得到否證後，大部分的失敗者（78.57%）只對自己原先的假設做互包性的修正，亦即可能只會進行弱重建，這樣的傾向對於迷思概念的根本重建可能會造成阻礙。例如兒童有「地球是扁平的」迷思概念（Vosniadou *et al.*, 1987），當教師點出兒童想法的錯誤並告知地球是圓的之後，兒童僅將其迷思概念作一弱重建–地球像一個盤子，是扁圓形的。所以教師在使用「異例法」進行概念改變教學時，除了在第二階段引入異例外，更應注重第三階段的調適，如何能讓學生由不同的角度（交疊性的假設而非互包性的假設）來解釋現象。如何利用「有限的雙假設測試法」要求學生在開始時想出兩個可能的解釋再進行測試，可能會是一個值得進一步探討的方式。

在本研究中，我們選擇「246作業」這個發現規則作業來研究學生的假設檢驗推理能力有兩個優點：第一、心理學家利用此作業進行假設檢驗推理的研究已經累積了大量的結果，提供我們豐富的基礎以進行比較並深入探討；第二、「246作業」所給的起始正例「2、4、6」只涉及數字概念，是參與者都具有的基本知識，並不會牽涉到大量結構性的知識或基模（schema），如此我們可

以暫時忽略先前知識對於推理的影響，而將焦點集中於推理的歷程上。未來若要更全面地瞭解人們測試與發現規則的機制以及概念修正的歷程，勢必將先前知識或基模的影響納入考量，尤其是概念改變教學中欲摒除的迷思概念常是學生由日常經驗中形成、可解釋大部分日常生活現象的概念或原則，這些先前概念常是很強、很難改變的。這些既有的信念或先前知識對於選擇測試策略、形成新的交疊性另有假設扮演什麼樣的角色是我們今後需要再加以探討的。

致　謝

本研究的完成，得到國科會（NSC 88-2511-S-002-011、NSC 89-2511-S-002-004）以及教育部卓越計畫（89-H-FA07-1-4-2）對第二作者的贊助，謹此致謝。我們也感謝胡志偉教授提供中文閱讀廣度測驗材料，姚開屏教授、翁儷禎教授在統計方法上的釋疑，以及黃庭和、林育聖、林逸鑫在資料分析上的協助。

附　註

1. Wason 稱前者為 enumerative thinking，後者為 eliminative thinking。
2. 由於我們是以先後或同時出現的兩個假設之間的案例集合是否為交疊性關係來定義交疊性另有假設，因此採用交疊性測試法時，一定是產生了交疊性假設。但反過來說，交疊性假設的產生則不一定是在使用交疊性測試法後。在其他測試法之後（大部分是得到否證後），受試者也有可能產生一個新的交疊性另有假設。
3. 本實驗的目的主要著重於資優生與一般大學生（包括所有學院）的比較，我們想看看資優生是否有較高的成功率、他們是否具有較多我們所發現的成功者的特色（產生較多交疊性假設和運用較佳策略）。至於是甚麼原因(例如 IQ、數理知識等等）造成資優生有這些特色，並不是本實驗的目的。但我們之後會討論工作記憶容量與知識的可能影響。
4. 如前述，除了互包與交疊性關係外，兩假設間也可能為錯開的關係（disjoint），但在「2 4 6 作業」中，兩假設的關係若為此，則其中必有一個為不可能之假設（non-plausible hypothesis），因為所有可能的假設都至少要包含「2、4、6」這個正例，故一定會有交集的部分。在本實驗中，並沒有受試者曾出現這種假設。
5. 比對兩評分者對每一位受試者每次測試策略與假設種類的分類，評分者間的意見相同率為分類一致總次數 / 總比對次數。
6. 我們分析一般生中不同的學習背景對正確率的影響，若將 29 名受試者分為文、法、社會科學、管理學院（10 人）以及理、工、醫、農學院（19 人）兩組，兩組中的正確率分別為 20 %和 21.2%。此結果顯示在「2 4 6 作業」中，數理知識背景並不會對推理的成功率造成影響，這可能是因為「2 4 6 作業」並不牽涉複雜難深的數理知識，一般大學生甚至中學生都可能具備。
7. 研究中採用 t 檢定前先經過 Levene 變異數同質性檢定，當變異數同質性假設不符合時，利用 Welch-Satterthwaite 的自由度修正(df　)來進行檢定。
8. 效果量 d 的計算是根據以下的公式：$d = t[(1/n_1)+(1/n_2)]^{5}$（Cortina, J. M. & Nouri, H., 2000），此處 t 為經 t 檢定所得之統計值，n_1 是第一組的人數，n_2 是第二組的人數。此公式用於不等格設計的狀況。
9. 由於畢業、出國等因素，使得其餘資優生有困難回來接受「工作記憶廣度測驗」。
10. 根據 Kyllonen & Christal (1990) 的設計原則自行設計。
11. 因為本實驗中對照組只有 4 個成功者，但因實驗二中對照組與實驗一中一般生組的實驗程

序與材料完全相同，且來自同一個受試者來源，為了能從統計中看出一些傾向，我們合併了兩組中的成功者（共 10 人）來與指導組中的成功者（共 18 人）進行比較。

參考文獻

1. 宋志雄、林曦和徐順益（1993）：探究國三學生酸與鹼的迷思概念並應用以發展教學診斷工具。科學教育月刊, **4**, 1-23。
2. 吳庭瑜、吳明撣、洪瑞雲 (1998)：合作學習、解釋及發問架構提示對歸納推理表現之影響。中華心理學刊, **40:2**, 117-136。
3. 洪振方 (1996)：科學知識重建的知識取向分析。高雄師大學報, **7**, 293-328。
4. 郭重吉和吳武雄（1989）：利用晤談方式探查國中學生重要物理概念的另有架構之研究(I)。國科會專題研究成果報告。彰化市：彰化師大。
5. 張川木（1995）：促進概念改變教學法(Ⅰ)。科學教育月刊, **185**, 21-27。
6. 張川木 (1996)：促進概念改變教學法(Ⅱ)。科學教育月刊, **186**, 10-18。
7. 陳貽照（1998）：影響中文多義詞在句中辨識歷程的因素：工作記憶容量或多義詞在語言中的比率。台北市：國立台灣大學心理學研究所碩士論文。
8. 連韻文（1998）：科學資優生哪裡資優？–以假設檢驗能力為例。資優教育二十五週年研討會論文專輯, 135-147。
9. 謝志仁和郭重吉（1993）：國中學生化學變化相關概念另有架構之研究。科學教育月刊, **4**, 25-51。
10. Baddeley, A. D. (1976). *The psychology of memory*. New York: Basic Books.
11. Carey, S. (1985). Are children fundamentally different kinds of thinkers and learners than adults? In S. Chipman et al. (Eds.), *Thinking and learning skills* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates.
12. Carey, S. (1986). Cognitive science and science education. *American Psychologist, 41*(10), 1123-1130.
13. Chang, C. M. (1993). Using a microcomputer-based laboratory in teaching selected concepts in mechanics. Unpublished ph. & Thesis University of heeds.
14. Cortina, J. M. & Nouri, H. (2000). *Effect size for ANOVA designs*. Thousand Oaks: SAGE Publications, Inc.
15. Evans, J. St. B.T. (1989). *Bias in human reasoning* : *causes and consequences*. London: Lawrence Erlbaum Associates.
16. Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 87*, 190-211.
17. Gorman, M. E. (1986). How the possibility of error affects falsification on a task that models scientific problem-solving. *British Journal of Psychology, 77*, 85-96.
18. Gorman, M. E., & Gorman M. E. (1984). A comparison of disconfirmatory, confirmatory and control strategies on Wason's 2-4-6 task. *The Quarterly Journal of Experimental Psychology, 36A*, 629-648.
19. Gorman, M. E., Gorman M. E., Latta R. M., & Cunningham G. (1984). How disconfirmatory, confirmatory and combined strategies affect group problem solving. *British Journal of Psychology, 75*, 65-79.
20. Gorman, M. E., Stafford, A., & Gorman M. E. (1987). Disconfirmation and dual hypotheses on a more difficult version of Wason's 2,4,6 task. *Quarterly Journal of Experimental Psychology*, 39A, 1-28.
21. Green, D. W. (1990). Confirmation bias, problem-

solving and cognitive models. In J.-P. Caverni, J.-M. Fabre, & M. Gonzalez (Eds.), *Cognitive biases*. Amsterdam: Elsevier.

22. Johnson-Laird, P. N. (1983). *Mental models*. Cambridge : Cambridge University Press.
23. Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 1, 122-149.
24. Kareev, Y., Halberstadt, N., & Shafir, D. (1993). Improving performance and increasing the use of non-positive testing in a rule-discovery task. *Quarterly Journal of Experimental Psychology, 46A*, 729-742.
25. Klahr, D., & Dunbar, K. (1988). Dual Space search during scientific reasoning. *Cognitive Science, 12*, 1-55.
26. Klayman, J., & Ha. Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211-228.
27. Klayman, J., & Ha. Y. W. (1989). Hypothesis testing in rule discovery: strategy, structure, and content. *Journal of Experimental Psychology* : *Learning, Memory and Cognition, 15*, 596-604.
28. Kuhn, T. S. (1970). *The structure of scientific revolutions*. (2nd ed.). Chicago: University of Chicago Press.
29. Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity ?! *Intelligence*, *14*, 389-433.
30. Lakatos, I. (1970). Falsification and the methodology of scientific research programmers. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-105). London: Cambridge University Press.
31. Laughlin, P. R., & Bonner, B. L. (1999). Collective Induction: Effects of multiple hypotheses and multiple evidence in two problem domains. *Journal of Personality and Social Psychology, Vol. 77*, No. 6, 1163-1172.
32. Laughlin, P. R., Bonner, B. L. & Altermatt T. W. (1998). Collective versus individual induction with single versus multiple hypotheses. *Journal of Personality and Social Psychology, Vol. 75*, No. 6, 1481-1489.
33. Lien, Y-W., & Lin, W-L. (in preparation). What is critical for discovering a rule : Investigating the role of falsification and alternative hypothesis.
34. Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effect of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098-2109.
35. Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, *29*, 85-95.
36. Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and short-comings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
37. Nussbaum, J., & Novak, S. (1982). Alternative frameworks, conceptual conflict and accom-modation: Toward a principled teaching strategy. *Instructional Science, 11*, 183-200.
38. Penner, D., & Klahr, D. (1996). When to trust the evidence: Further investigations of the effects of system error on the Wason 2-4-6 task. *Memory & Cognition, 24*(5), 655-668.
39. Platt, J. R. (1964). Strong inference. *Science, 146*, 347-353.
40. Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books.
41. Ross, L., & Lepper, M. R. (1980). The perseverance of beliefs: Empirical and normative considera-tions. In R. A. Shweder (Ed), *Fallible*

judgment in behavioral research: New directions for methodology of social and behavioral science (Vol. 4, pp. 17-36). San Francisco: Jossey-Bass.

42. Rossi, S., Caverni, J. P., & Girotto, V. (2001). Hypothesis testing in a rule discovery problem: When a focused procedure is effective. *The Quarterly Journal of Experimental Psychology, 54A*(1), 263-267.
43. Snyder, M. (1980). Seek and you shall find : Testing hypotheses about other people. In E. T. Higgins, C. P. Heiman, & M. P. Zanna (Eds.), *Social Cognition: The Ontario Symposium on personality and social psychology* (pp. 277-303). Hillsdale, NJ: Erlbaum.
44. Snyder, M., & Swann, W. B., Jr. (1978). Snyder, M., & Campbell, B. H. (1980). Testing hypotheses about other people: The role of the hypothesis. *Personality and Social Psychology Bulletin, 6*, 421-426.
45. Hypothesis-testing in social interaction. *Journal of Personality and Social Psychology, 36,* 1202-1212.
46. Torrance, E. P. (1989). The Nature of creativity as manifest in its testing. In Sternberg, R. B. (Ed.), *The nature of creativity* : *Contemporary psychological perspectives*. Cambridge : Cambridge University Press.
47. Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A., & Arkkelin, D. L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology*,*32*,109-123.
48. Vosniadous, S., & Brewer, W. F. (1987). Theories of knowledge restructuring in develop-ment. *Review of Educational Research, 57,* 51-67.
49. Wharton, C. M., Cheng, P. W., & Wickens, T. D. (1993). Hypothesis-testing strategies: why two goals are better than one. *Quarterly Journal of Experimental Psychology, 46A,* 743-758.
50. Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129-140.
51. Wason, P. C. (1966). Reasoning.In B. M. Foss (Ed.), *New horizons in psychology* (pp.135-151). Harmondsworth, Middlesex, England : Penguin.

附錄一　測試策略分類範例

以下我們以第一個測試做為例子說明四種測試策略分類的準則，一開始給予符合規則的例子〔2, 4, 6〕，然後受試者寫下他認為可能的假設，在此我們以最多受試者寫下的假設—「偶數」為例來說明接下來的測試如何分類。表格左邊是分類結果，右邊是一個範例，代表受試者所寫下該次測試的例子與理由，是據以分類的主要訊息。

正例法	**測試的例子：**　「6、10、24」 **測試的理由：**　看規則是不是「偶數」

反例法	測試的例子：　「3、6、9」 測試的理由：　看規則是不是「偶數」
交疊性測試法	測試的例子：　「3、6、9」 測試的理由：　規則有沒有可能是「等差數列」 說明：在「偶數」之外，又想到「等差數列」，兩假設之間的關係為交疊性。所用的測試例子可以區辨這兩個假設何者為真。
互包性測試法	測試的例子：　「6、10、24」 測試的理由：　規則有沒有可能是「相差2的偶數」 說明：在「偶數」之外，又想到「相差2的偶數」，兩假設的關係為互包性。所用的測試例子可以區辨這兩個假設何者為真。

附錄二　工作記憶廣度測驗範例

閱讀廣度測驗：以下我們舉一個二句題、一個三句題作為範例，表格左邊的數字是電腦螢幕呈現的順序，表格右邊則為螢幕上每次出現的中文句子，句末的雙字詞在螢幕上是以紅色呈現，受試者朗讀完一句後，接著朗讀下一個出現的句子，直到要求回憶的指導語出現。四句、五句以及六句題所朗讀的句子分別為4、5、6句。受試者若通過5題二句題中的3題，則繼續做三句題，否則即終止；若通過5題三句題中的3題，則可再做四句題，其餘以此類推。

閱讀廣度測驗範例一、（二句題）

1	他一方面要安撫家人的情緒，另一方面又要面對外界的壓力。
2	在政治結構轉變的過程中，我們應該小心處理敏感的憲政改革問題。

3	請回憶一在答案卷上順序寫下剛才出現的紅字詞

閱讀廣度測驗範例二、（三句題）

1	經過兩天不眠不休的工作，他累得連飯都沒吃就睡覺了。
2	這年頭夠資格的人真是太多了，多得讓人不能相信。
3	這麼久了，窗外的毛竹依舊那麼青翠，窗裡的人卻蒼老了。
4	請回憶一在答案卷上順序寫下剛才出現的紅字詞

數字工作記憶測驗：以下我們舉出一個數字工作記憶測驗的範例，表格上方的數字是電腦螢幕呈現的順序，1至3螢幕呈現2秒鐘後消失，接著受試者按鍵回答4至6的問題。

1	2	3	4	5	6
A＝2B-3	B＝C/4	C＝8	B＝？	A＝？	C＝？

How to Discover a Rule? Investigating the Advantages of the Alternative Hypothesis and Dual-Hypothesis Testing Strategy

Wei-Lun Lin and Yunn-Wen Lien

Department of Psychology, National Taiwan University

Abstract

Our research intended to find out what strategy makes a reasoner perform better in a rule-discovery task. In Experiment 1, two groups of 29 college students were tested with the 2 4 6 task, a popular rule-discovery task. One group was composed of undergraduates who were judged by experts and professors as top students in nature science and life science when they were high school students. These students subsequently were enrolled in relevant programs in National Taiwan University. The results showed that these students performed better on the task than did the contrast group, composed of students from the same university. Particularly, compared with the contrast group, these students generated more alternative hypotheses as well as applying more dual-hypothesis testing strategy. They, however, did not use negative-test strategy more often than the contrast group. We argue the way how a reasoner revised his/her hypothesis based on new evidence is crucial for the success on the 2 4 6 task. In addition, the potential advantages and constraints of using a dual-hypothesis testing strategy were discussed, as well as its relation with the reasoners' working memory. In Experiment 2, a group of university students were instructed on the limited dual-hypothesis testing developed by the researchers. Their performance in the rule-discovery task significantly improved when compared with the contrast group.

Key words: Working Memory, Alternative Hypothesis, Science-Talented Student, Hypothesis Testing, Conceptual Change.