

行政院國家科學委員會專題研究計畫期末報告

科學資優生語文能力之探討-
科學資優生語文推理能力分析

計畫類別：整合型計畫

計畫編號：NSC 89-2511-S-002-004

執行期間：88年8月1日至89年7月31日

整合型計畫：總計畫主持人：洪蘭

子計畫主持人：連韻文

執行單位：臺大心理系

中華民國 89 年 10 月

摘要

本研究以「246問題」作為探討假設檢驗推理能力的標準作業，針對和科學研究潛能有重要相關的假設檢驗能力，做一深入探討。在前兩年的計畫中，我們先以一般大學生為樣本，發現影響假設檢驗推理能力成敗的關鍵在於是否能夠產生替代性的假設（alternative hypothesis），是否使用反例法並不重要，否證量的高低也與成功率無關。我們比較科學資優生與一般大學生，發現資優生產生了較多的替代性假設，並較常使用雙假設測試，在此作業上得到較高的成功率。考量資優生的表現，我們成功地以有限的雙重假設的設計，使得受試者產生更多的替代性假設，而得到較高的成功率。此結果可作為推理教學的參考。

今年度的研究，進一步探討甚麼因素影響替代性假設的產生，我們操弄工作記憶的負擔，發現減輕受試者的工作記憶負擔，能使其產生較多的替代性假設，但不會影響另一種新假設（原假設範圍的放大或縮小）的產生，並使用較有效率的替代性測試策略，而提高推理的正確率。並以兩種工作記憶廣度測驗，篩選出工作記憶廣度高者與低者，比較他們在檢驗推理作業的表現。如所預測，廣度高者，產生較多的替代性假設，但不是範圍修正性假設。

關鍵詞：工作記憶、否證、替代性假設、假設檢驗推理、科學資優生

Abstract

This research focuses on the reasoning ability regarding hypothesis testing, which is critical for conducting scientific research. In the first two years of the project, we found that whether subjects could successfully discover a rule predetermined by the experimenter depended on the number of alternative hypothesis they generated, and the use of negative-test strategy was not critical. Opposed to what other researchers speculated, the amount of disconfirming evidence is not critical, either. We also found that the performance could be greatly improved by forcing subjects to generate two hypotheses at a time in each of the first few test trials.

Based on the finding, we compared the difference between normal college students and those who were selected by Ministry of Education as scientific-talented students. The results showed that scientific-talented students performed better on '246 task', and they generated more alternative hypotheses. In addition, scientific-talented students used less positive-test strategy. That probably means that they have better chance to escape from the so-called 'confirmation bias'.

In the third year of the project, we further investigate what makes a person produce more alternative hypothesis than others. By manipulating subjects' working memory load as well as measuring their working memory span directly, as predicted, we found that the less working memory available, the fewer alternative hypotheses could be generated, but the generation of boundary-adjusted hypotheses is not affected.

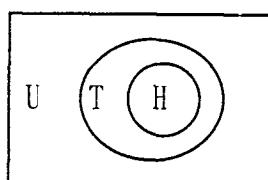
Keywords : hypothesis testing, working memory, falsification, rule discovery

一個好的科學家，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。著名的研究創造力的學者 Torrance (1985) 也認為假設的檢驗與產生也是創造性思考的環節之一，而說道「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」。我們可以說，假設檢驗推理是人類一個基本且重要的認知能力，對於科學研究來說，更是不可或缺的。但是過去的心理學研究發現人們在一些簡單的規則發現作業中（例如 Wason 的 246 作業），傾向於使用正例來測試自己的假設，亦即測試符合自己假設的例子，而較少測試不符合自己假設的例子。甚至如果用正例法得到一些肯定的證據後，就會很有信心認為他的假設是對的，而不會積極地尋求否證 (e.g., Wason, 1960, 1968)。這些現象也在科學家、統計學家等身上發現 (e.g., Einhorn & Hogarth)。使用正例測試法在一般情形下，不失為一種有效得到否證的策略，但在某些情況下，過度使用正例法卻會沒有得到否證的機會 (Clayman & Ha, 1987, 1989)。根據 Popper (1959) 的否證理論，得到否證才是一個科學理論或假設得以修正的要素，得到支持的證據並無法證明假設或理論為真，因此人們這種尋求正例的傾向常被心理學家視為是一種「肯證偏誤」(confirmation bias)。因此，瞭解與評估學生在假設檢驗上的策略與能力十分重要，不但可以作為挑選有科學潛力的科學資優生或是創造力評估的一項參考，也可對如何增進學生這方面的能力有所啟發。

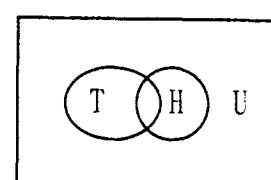
246 作業的研究回顧

Wason(1960)所設計的「246 作業」及其變型可被視為日常生活或科學研究規則發現 (rule discovery) 的一個簡化作業，也是研究假設檢驗推理最常用的作業，我們這系列的研究採用此作業。在此作業中，實驗者預設一個規則（例如遞增數列），受試者只被告知一個符合此規則的例子（例如 2、4、6），據此猜測實驗者的預設規則。受試者在隨後的每次測試中可提出一組例子（例如 4、6、8）來測試其猜測（亦即受試者對預設規則的假設，例如「偶數」），在每次測試之後，實驗者都會依照此例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或在測試了規定的測試次數後，受試者要猜出

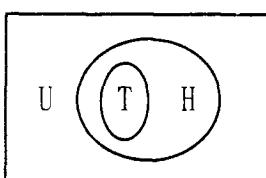
預定的法則。Wason 在此作業中發現，79%的受試者很有信心的停止測試，但是答案卻是不正確的。受試者有找尋支持自己原有假設的例子或訊息（即正例，見圖一）的傾向（例如以 4、6、8 來測試「偶數」這個假設對不對），較不會尋求有可能不支持自己假設的例子（即反例，例如 1、3、6），以至於不易發現自己假設的錯誤。Klayman 和 Ha (1987) 曾提出初始假設 (H) 與預設規則 (T) 之間的關係可能有 4 種（見圖一），分別為包含 (embedded)、交會 (overlapping)、圍繞 (surrounding) 與錯開 (disjoint)，並指出在不知初始假設與預設規則間的關係之下，尋求正例不失為一種有效的策略，但不可否認的，在「246 作業」中（受試者的初始假設常與預設規則是包含的關係），正例法並不有效（成功率多在 40% 以下）。



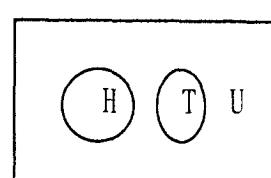
內含 (embedded)



交會 (overlapping)



圍繞 (surrounding)



錯開 (disjoint)

圖一、初始假設 (H) 與預設規則 (T) 之 4 種可能關係圖 (U: universal set)

為了改善在檢驗假設作業上表現，過去許多研究者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，或是改變作業的程序，以直接或間接增加否證的訊息來提高作業的成功率 (e.g., Tweney, 1977, 1980; Gorman, 1984, 1986, 1987; Klayman & Ha, 1989; Green, 1990)，但並沒有一致正面的效果。

我們的分析發現，過去研究者對於「否證」的定義並不一致，有些受試者認為「否證」即是反例，即認為多使用反例測試法，便可得到較多的否證。

(Tweney, 1977, 1980 ; Gorman, 1984, 1986, 1987 ; Klayman & Ha, 1989 ; Green, 1990 ; Halberstadt & Kareev, 1992, 1993 ; Penner & Klahr, 1996.)；另一些實驗者將「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) (Gorman, 1984, 1986, 1987 ; etc.)；至於較符合 Popper (1959) 應是認為「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。Klayman & Ha (1987) 雖提出上述論述，卻無人以此作為分析的依據。

本研究的目的即在於根植於過去的研究與問題，探討否證與其他因素對假設檢驗作業成功的影響。並比較現有的科學資優生（通過教育部甄選保送各大學基礎自然學科者）與一般大學生在假設產生與檢驗方面的作業上所採取的策略，是否科學資優生會使用不同或較佳的策略？是甚麼因素導致有些人可以使用較佳的策略，有些不能？此能力是否可因教導而改善？以作為科學或推理教育的參考。以下先略述前兩年的研究結果，以導入今年的研究成果。

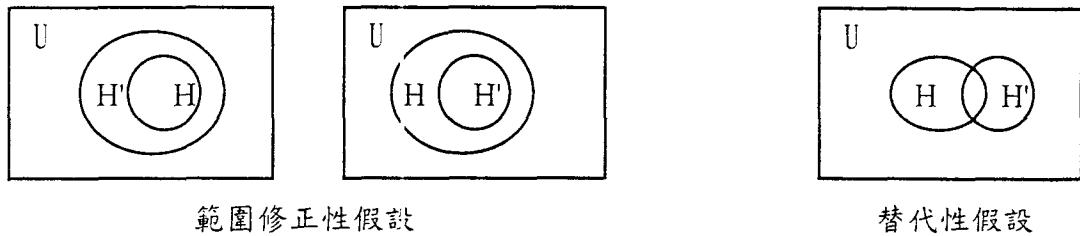
第一年與第二年研究成果

在前兩年的計畫中，我們以一般大學生為對象，探討在假設檢驗作業中成功者與失敗者在策略與假設產生種類上有無差異，以及分析否證多寡的影響，並以此為基礎，評估科學資優生在這方面的表現，並試圖找出可以改善此作業表現的策略。本研究中的資優生定義為 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」的考核通過而進入台灣大學基礎科學的科系就讀的學生。

成功發現規則者與失敗者的區別

我們的研究指出，得到否證雖有助於刪除不正確的假設，也能促進新假設的產生，但這新假設常常只是原有假設範圍的擴大或縮小（新舊假設間為包含關係），而不是形成新角度的假設（新舊假設之間只有部份重疊）。正如科學哲學家 Lakatos (1968) 繼 Popper 之後提出的「精緻否證論」認為，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。過去研究將所有的新假設都視為替代性假設，將此兩種性質的假設混為一談。

一談，但是，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的假設（也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），需要從新的角度來看，因此若只產生範圍修正性假設，並不會促進規則的發現，得到否證訊息卻失敗的受試者，可能就是陷入範圍修正的循環中，所以有必要對範圍修正性假設與真正的替代性假設作一區分（見圖二）。研究結果顯示，區分範圍修正性假設與替代性假設之後，成功者在替代性假設的個數上顯著多於失敗者，但在範圍修正性假設或假設總數上並沒有差異。



圖二、替代性假設與範圍修正性假設之圖示

推理者在假設檢驗的推論過程中也不會只採用上述提及的正例法或反例法，每次只針對一個假設來測試。有些推理者會根據已知同時形成兩個假設來進行測試（Klayman & Ha 稱之為替代性測試）。我們進一步根據所產生的兩個假設之間的關係，將雙假設測試分為替代測試法（同時測試兩個有部份疊合關係的假設）與範圍修正測試法（同時測試兩個有包含關係的假設）。使用替代測試法即是在測試兩個替代性假設。我們發現成功者比失敗者使用更多的替代測試法，但在範圍修正法、正例與反例法的使用上並無差異。此外我們分析了前述三種不同定義的否證，如我們所預測的，成功與失敗者之間並未有差異。

科學資優生與一般生的區別

科學資優生的行為較類似上述成功者的特色，而他們也的確在正確率上顯著高於一般生。如我們所預測的，資優生的替代性假設較一般生多，但在範圍修正性假設上與一般大學生沒有差異。在測試策略方面，資優生較少使用正例法，而較多使用替代法與範圍修正法，因而具區辨性的雙假設測試的使用也顯著多於一般生。至於反例法的使用，則與一般生並無不同。兩組在否證量上並沒有差異。

資優生為何能夠主動的採用較多雙假設測試，而不僅只用單一假設去進行測試呢？由於使用雙假設測試策略必然會較使用單一假設測試策略需要耗費更多的工作記憶資源，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，我們認為資優生會常採用這種有效但須較多工作記憶的策略，除了可能是學習到較佳策略外，也可能是因為他們具有較大的工作記憶廣度，故較一般學生更能主動使用。

如何提高假設檢驗作業的表現？

從資優生的實驗結果中可以看出，使用雙重假設測試策略應該有它的正面效果，但是在過去的研究中，Tweney (1980) 曾要求受試者每次測試兩個假設，這樣的操弄卻造成了反效果，受試者的表現顯著降低。我們認為，這是因為 Tweney 並未考慮工作記憶負擔的影響，要求受試者從頭至尾都同時測試兩個假設，造成太大的工作記憶負荷，受試者反而亂了頭緒。因此我們根據資優生特別的思考策略，同時考慮工作記憶的影響，改善 Tweney 作業的缺點，設計「有限的雙假設測試法」，只要求受試者在前四個測試中，每次都要測兩個假設（但並未規定假設之間的關係），其餘測試則無規定，結果大幅度地改善他們在假設檢驗方面的表現。

工作記憶廣度對替代性假設產生的影響

心理學研究顯示工作記憶廣度大，可同時考慮多個不同向度的線索及假設，而有助於推理，因此我們更進一步探討過工作記憶廣度與替代性假設產生的關係。我們採用心理學在探討工作記憶時所採用的「雙作業」實驗派典，發現若操弄受試者在做 246 作業時工作記憶的負擔，則工作記憶減輕組的正確率顯著的高於加重負荷組。更重要的是減輕組所產生的替代性假設顯著高於加重負荷組與標準組，範圍修正性假設則沒有差異，測試策略亦無不同。此結果顯示替代性假設的產生會受工作記憶影響。

在今年度裡，我們採用另一種方式來探討工作記憶廣度與替代性假設產生的關係，我們設計兩個以電腦呈現的工作記憶廣度測驗來篩選高、低廣度的受試者，再令他們進行「246 作業」，以更直接的方式再次檢證工作記憶廣度對替代性假設的產生的影響。同時也檢測科學資優生在工作記憶廣度測驗的表現，來印證我們的解釋。

方法

受試者 105 名台灣大學學生應課程要求來參加實驗，實驗中分別進行兩個工作記憶廣度測驗（順序隨機），挑出 18 名高分組與 19 名低分組進行 246 作業。

材料及程序

工作記憶廣度測驗：我們編製兩種工作記憶廣度測驗，並將其電腦化。

(1) 閱讀廣度測驗（根據胡志偉（1996）的題本）- 包括 100 句中文句子，五種難度，分數由 1 分至 6 分，句子由電腦呈現。受試者由最簡單的兩句題開始，每朗讀一個句子，同時要記下句中用紅字呈現的詞彙，唸完兩個句子之後，將所記下的兩個詞彙順序回憶出，完成 5 個兩句題之後，再向上進行 3、4、5、6 句題，每一程度也都有 5 題。5 題中若能正確回憶出 3 題以上，則算是成功達到此一程度，若 5 題中只能正確回憶出兩題，則此一程度只能得 0.5 分。例如有一個受試者 A 在 3 句題中答對 3 題，他的得分則為 3 分；另有一受試者 B 在 3 句題中答對 3 題，4 句題中答對 2 題，他的得分則為 3.5 分。

(2) 數字工作記憶測驗（根據 Kyllonen & Christal (1990) 的設計原則設計）- 共 25 題電腦測驗，五種難度，分數由 0 至 75 分，每一題中要受試者算出 A、B、C 的數值。測驗進行時，電腦螢幕首先會呈現一個代數式，說明 A 與 B 的關係，但這個代數式 2 秒鐘就會消失，接著出現第二個代數式，說明 B 與 C 的關係，這個代數式 2 秒鐘也會消失，再來出現 C 的數值，受試者必須記住兩個代數式和 C 的數值，經由心算來解出 A、B、C 的數值，每答對一題得一分。此兩種測驗的例題見下表一。

246 作業：藉由工作記憶測驗挑選出低分組與高分組，再進行「246 作業」，程序與之前實驗相同（請參照 88 年度的成果報告）。

表一、工作記憶廣度測驗範例

閱讀廣度測驗範例（兩句題）

他一方面要安撫家人的情緒，另一方面又要面對外界的。



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。



請回憶---在答案卷上順序寫下剛才出現的紅字詞

數字工作記憶測驗範例

中等難度(六個等式依序出現)

(1) $A = 2B - 3$

(2) $B = C/4$

(3) $C = 8$

(4) $B = ?$

(5) $A = ?$

(6) $C = ?$

較高難度

$A - 1 = 2B + 2$

$B + 3 = 3C - 5$

$C = 3$

$B = ?$

$A = ?$

$C = ?$

結果

區分標準 高分組：在兩測驗中任一測驗的得分在前 25% 以上，同時另一測驗的得分也在全體平均以上(數學 66 分以上[佔全體 $27/105=25.71\%$]且 Reading Span 3.5 以上，或是 Reading Span 4 以上[佔全體 $25/105=23.81\%$]且數學 57 以上)；低分組：在兩測驗中任一測驗的得分在後 25% 以下，同時另一測驗的得分也在全體平均以下 (數學 4" 以下[佔全體 $25/105=23.81\%$]且 Reading Span 3 以下，或是 Reading Span 2.5 以下[佔全體 $22/105=20.95\%$]且數學 56 以下)。由 105 名受試者中挑選出 18 名高分組與 19 名低分組。高分組與低分組與全體的平均得分見下表二。

表二、工作記憶測驗平均得分

	高分組	低分組	全體
數學題	65.56	45.32	56.26
Reading Span	4.39	2.42	3.31

「246 作業」正確率 高分組正確率 61.1%，低分組正確率 10.5%，兩組達顯著差異 ($\chi^2=10.38$, $p < 0.01$)。

假設種類、測試策略與否證量 如下表三所示，在假設種類方面，高分組替代性假設的產生顯著高於低分組 ($t=2.71$, $p < 0.01$)，這使得高分組在假設的總數上也多於低分組 ($t=2.02$, $p=0.05$)，但兩組在範圍修正性假設的個數上沒有差異。在測試策略方面，低分組使用較多的正例法 ($t=-2.13$, $p < 0.05$)，高分組使用較多的反例法 ($t=2.27$, $p < 0.05$)，在其他的測試策略使用上沒有差異。在否證量方面，雖然高分組使用較多的反例法及較少的正例法，但在所得到的否證量上，高低分組並沒有差異。

表三、高低分組的假設種類、測試策略與否證量

平均個數(個/12 嘗試)		高分組 (18人)	低分組 (19人)
測試策略	正例法	3.61	5.63 *
	反例法	2.33	1.16 *
	替代法	0.56	0.32
	範圍修正法	5.11	4.58
假設種類	替代性假設	1.39	0.58 **
	範圍修正性假設	6.39	5.42
假設總數		8.78	7.00 *
肯證數		4.33	5.84
否證數		1.61	0.95

* $p < .05$ ** $p < .01$

討論

這個實驗結果再次支持工作記憶會影響替代性假設的產生，而和範圍修正假設的產生無關，這也支持我們將新假設區分為這兩種是有意義的。

一個和我們過去實驗結果不同的發現是，和低廣度組比較，高廣度組使用較多的正例法，以及較少的反例法。過去的研究會將這種傾向解釋為廣度高組有可能獲得較多的否證，但我們的結果顯示高低工作記憶廣度者在所得到的肯

證量與否證量上無差異，儘管他們在作業的正確率上有極大的差異。反例法和正例法都有可能得到否證，也各有不能得到否證的情形(Clayman & Ha, 1987)，我們的結果顯示即便是在過去學者認為有利於反例法的情況下(標準 246 作業的設計，會使得使用正例法常常無法得到否證)，反例法的使用頻率仍與整體否證量無關，這是因為隨著受試者假設的改變，會改變假設和正確規則間的相互關係，而使正例與反例各有有利的時候。我們認為用另一個角度來看正例與反例的差異，可能較有意義。由結果可知，高廣度組在使用正例與反例法上較均衡，但低廣度組，有較大比例使用正例法。在不知道正確規則的情況下，前者無寧是一種較不會陷入「肯證偏誤」的策略。

結論

綜合我們這個三年的研究，有以下幾點新的發現，第一，過去的研究皆注重於探討假設檢定測試策略的重要性，尤其是關注在反例法的使用不足，因而探討各種可提高反例法使用的指導以提高正確率，但效果並不一致。我們的研究將焦點轉移至假設的產生上，並進一步區分所產生的假設，認為替代性假設的產生才是假設檢驗推理成敗的關鍵，使用反例測試策略並不重要。

第二，過去大家認為否證的多寡關係者發現正確規則的機率，但並未有人真正分析受試者所得的否證量。我們也發現過去研究對否證定義不一致之處，並分別分析，發現否證量並非區辨成功找到規則與否的關鍵因素。

第三，雖然 Klayman & Ha 在 1989 年曾指出替代性測試比反例法重要，但是他們並未清楚定義替代性測試，事實上，Klayman & Ha 的替代性測試包含了我們我們所說的替代性測試與範圍修正性測試，我們的研究顯示，這種區別是必要的。

第四，我們由資優生的實驗結果中也同樣看出，資優生產生了較多的替代性假設，而得到較高的成功率。另外，我們發現資優生在假設檢驗的過程中較一般學生傾向使用雙假設測試策略，這樣的傾向更有助於資優生產生替代性的假設。由於使用雙假設測試策略較使用單假設測試策略需要耗費更多的工作記憶資源，所以我們認為資優生能主動頻繁地採用此一策略，顯示出資優生可能具有較豐富的工作記憶資源。

第五，我們也發現不論是間接的操弄工作記憶的負擔或是設計測驗，直接測量工作記憶廣度，在工作記憶資源不足或少的情況下，受試較傾向採用單假設的測試策略，較不會主動採用雙假設測試法。且如所預測，會產生較少的替代性假設，但不影響範圍修正性假設。

第六，根據資優生與一般生的不同表現，我們在考量了增進假設檢驗表現的關鍵因素之後，設計有限的雙重測試法，成功地提升一般大學生的表現。我們發現在不超過工作記憶負擔的情形下，同時測試兩個假設，能有效的提高替代性測試策略的使用，因而會提高替代性假設的個數，增加正確率。這樣的結果有助於以後在推理教學上的應用，期能提升學生在思考推理方面的普遍素質。

參考文獻

- Baddeley, A. D. (1976). *The psychology of memory.* New York: Basic Books.
- Evans, J. St. B.T. (1989). *Bias in human reasoning: causes and consequences.* London:Lawrence Erlbaum Associates.
- Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective.*Psychological Review*,87,190-211.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory.*Psychological Review*,99,1,122-149.
- 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(4),729-742.
- Klayman, J., & Ha, Y.W.(1987). Confirmation,disconfirmation, and information in hypothesis testing.*Psychological Review*,94,211-228.
- Klayman, J., & Ha, Y.W. (1989). Hypothesis testing in rule discovery: strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15,4,596-604.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*,14,389-433.
- 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.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129-140.

Wason, P. C. (1966). Reasoning. In B.M.Foss(Ed.), *New horizons in psychology*(pp.135-151). Harmondsworth, Middlesex, England: Penguin.