數字系列完成測驗試題認知成分分析之研究

林世華

國立台灣師範大學
教育心理與輔導學系

葉嘉惠

台北市
石牌國中

本研究之目的在結合認知成分分析與心理計量分析的觀點來了解影響數字系列完成作業試題難度的試題成分，以及分析在數字系列完成測驗中所測量之心理構念是否因爲數列規則破壞與否而有所不同。研究題材乃採用青年智力測驗中的數學測驗 446 名高二學生預試資料；在分析方法方面，同時採用多元逐步迴歸方法以及線性對數潛在特質模式來了解試題成分對於試題難度的影響。本研究之主要發現如下：㈠編製數字系列完成測驗時，加入不規則數列不會造成試題反應維度的改變。㈡影響數字系列完成測驗試題難度的試題成分爲「是否爲移動數列」、「週期長度」與「字串間是否有假造相同的數字」等三類；此三類試題成分對於試題難度皆具有正向的預測力。㈢試題成分之間的交互作用也是試題難度變異的來源之一。㈣利用多元逐步迴歸方法以及線性對數潛在特質模式進行試題成分分析所得結果頗爲一致。

關鍵詞：數字系列完成測驗、認知成分分析

緒論

十九世紀以來心理學家對於智力的研究方法已有重大的改變，按 Sternberg（1978）的分類，智力的研究約可分成傳統心理計量分析取向與認知成分分析取向兩種，傳統心理計量分析取向主要是延續差異心理學的觀點，藉由比較不同群體在標準化測驗上的差異表現來區別何謂高低智力；認知成分分析取向則係經由分析受試者在標準化測驗上的認知歷程來了解人類智力的本質。與之相對應的是不同的智力觀，其中心理計量分析取向多同意智力是由不同的因素所組成，而認知成分分析取向則認爲智力是表現於個體處理認知作業時所涉及的認知歷程。

認知心理學者對於智力的研究爲傳統心理計量研究帶來相當的衝擊，而其挑戰多集中於傳統智力測驗分數的意義以及測驗編製的方法。傳統智力測驗的編製多以古典測驗理論爲基礎，以測驗總分代表人類的智力，如此的方式固然能用以區辨不同個體的表現

，卻忽略了個體在智力測驗上的認知歷程。此外，傳統智力測驗的編製方法常被批評爲黑箱作業，測驗的編製並非事前依據特定的理論模式，而是藉由事後的統計分析方法來證實測驗的效度。

面對認知心理學者如此的挑戰，心理計量學者也有下列的回應。首先，項目反應理論(Item Response Theory)的發展使得測驗分數的解釋從古典測驗理論(Classic Test Theory) 中測驗相依（test dependent）的意義轉變成爲測驗獨立（test independent）的意義。其次，試題作答理論後期發展出的諸多模式也使得認知測驗的編製與認知變項的結合成爲可能。然而這樣的回應無異於顯示心理計量學者對於認知心理學者的挑戰束手無策：測驗分數的解釋雖能擺脫測驗相依的意義，但卻仍無法藉由測驗的分數了解受試者在測驗情境中所發生的所有認知歷程。此外，試題作答理論後期陸續發展的模式雖然使得測驗編製與認知變項的結合成爲可能，但認知變項的處理從來不是心理計量學者的專長。

不過心理計量學與認知心理學兩者多年來對於智力測驗的研究卻也爲整合兩者理論與方法提供了發展的基礎。認知心理學強調解題歷程，視題目爲其所包含的認知成分所組成，因此測量的基本單位不是題目，而是各個題目所涉及的認知成分，並且測量的單位若越詳盡明確，則各個試題所能提供的訊息量也將越豐富。而試題作答理論在一九七〇年代初期所發展出來的線性對數潛在特質模式（Linear Logistic Latent Trait Model, LLTM）除了依循試題作答理論以個別試題爲基本分析單位的傳統外，並且嘗試將試題依其解題步驟或是成分作線性分解，如此在估計試題的難度時，能夠同時考慮試題所涉及的試題成分或解題步驟，而且個別試題難度的解釋以及整體測驗試題難度的變異來源皆能事前加以掌握。

然而認知心理學對於智力測驗的研究也存在著下列的問題。認知成分分析多是直接採用現有的智力測驗或認知作業爲分析題材，再據之建立認知模式與考驗模式的適合度。如此的研究結果固然能提供智力測驗實質的心理學意義，但是認知成分的界定向來被批評流於主觀與含糊不清。此外，由於認知成分取向多直接採用現成的智力測驗或認知作業爲分析題材，以某一特定的測驗爲主體所建構的認知模式是否能概化至其他測驗，也是其招人質疑之處。認知成分分析的限制也促使近年來 LLTM 分析的主體從單純的認知成分，漸而包含可觀察、可測驗的試題成分。

基於前述動機，研究者欲結合認知成分分析取向以及心理計量分析取向，以系列完成作業爲題材，探討影響數字系列試題難度的試題成分；此外並探究藉以偵測規則系列的心理能力是否與偵測不規則系列的心理能力爲同一維度。具體言之，本研究所欲探討之問題爲： 1.偵測規則數列與不規則數列所涉及的心理構念是否相同？ 2.影響數字系列完成作業試題難度的試題成分爲何？ 3.數字系列完成作業試題成分與試題難度的關係式爲何？

文獻探討

系列完成作業是心理計量學者用以驗證普通因素(g)智力的測驗，並且也是認知心理學家參與智力研究採用的認知作業。認知心理學雖然係自一九七〇年代之後才興起智力測驗的研究，不過 Simon 與 Kotovsky 卻早在一九六三年便針對 Thurstone 基本心理能力測驗（PMAT）中的文字系列完成測驗進行認知成分分析。該篇研究指出，受試者在解決文字系列完成作業時，主要是藉由辨識各文字之間是否存有相同（Identity，例如"aaaa"）、下一個（Next，例如"abcd"）、上一個（ Backward，例如"gfed"）三種關係來辨識系列的規則。而受試者解題時其內在訊息處理歷程可分成「關係偵測」、「發現週期長度」、「完成組型描述」、「探測答案」等四個階段（Simon & Kotovsky, 1963；Kotovsky & Simon, 1973）。其後 LeFevre 與 Bisanz（1986）利用受試者在規則數列（例如"1 2 3 4 5 6 7 8 9 10"）、無效數列（例如"1 2 3 4 5 6 7 7 9 10"）與不規則數列（例如"1 2 3 4 5 11 7 8 9 10"）三者間的差異表現澄清了在 Simon 與 Kotovsky 所提出的認知模式「關係偵測」階段尚可再細分爲「記憶數列的再認」、「計算」、以及「檢視」等三個認知歷程。

長期以來心理計量分析取向對於系列完成作業的研究多著重關聯心理學與差異心理學的觀點（Sternberg, 1981, p.3），易言之，其著眼點多在於分析不同群體在標準化系列完成作業上表現的差異。然而近年來心理計量分析取向對於系列完成作業的研究也逐漸朝向了解不同試題內容對於受試者差異表現的影響，此類的研究尤以 Holzman, Pellegrino, 與 Glaser（1983）以及 Butterfield, Nielsen, Tangen, 與 Richardson（1985）的研究爲代表：Holzman 等人（1983）的研究主要是藉由自編數字系列完成測驗來驗證 Simon 與 Kotovsky 所提出的認知模式，以及探討影響數字系列完成作業試題難度的因素爲何。而 Butterfield 等人（1985）的研究更是結合了先前 Simon 與 Kotovsky 以及 Holzman 等人的研究結果，驗證「表徵文字系列的知識層級」、「計算次數」以及「關係複雜度」等三個變項是否足以預測系列完成作業的難度。

認知成分分析取向以及心理計量分析取向對於系列完成作業的研究似乎已臻完美，然而再深究各篇研究之後可發現尚有不足之處。首先，Simon 與 Kotovsky 所提出之認知模式雖然已成爲後續研究的準則，然而 LeFevre 與 Bisanz 的研究也指出 Simon 與 Kotovsky 的解題模式尚可再細分出其他的認知成分，這顯示該模式可能尚有其他未界定清楚的成分待進一步澄清。而 LeFevre 與 Bisanz 雖然藉由受試者在規則數列、無效數列以及不規則數列三者的表現說明了 Simon 與 Kotovsky 的解題模式的不足，但這樣的試題設計卻也是種大膽的嚐試。傳統因素分析研究雖顯示系列完成作業可用以驗證普通因素(g)智力，但傳統系列完成作業多爲規則系列，一旦將其規則破壞成爲無效系列或不規則系列之後，藉以偵測無效系列與不規則系列的心理能力是否仍與偵測規則系列的心理能力爲同一

維度，仍待進一步驗證。

Holzman 等人對於數字系列完成作業的研究乃建基於 Simon 與 Kotovsky 所提出的認知模式，不過該篇研究也顯示一個可能混淆的問題：Holzman 等人的研究顯示數字系列的數值大小以及運算種類會影響受試者解題表現，如此用以區辯不同受試的解題表現者究竟是普通因素智力亦或是過往數學學習成就，是其可能混淆之處。而 Butterfield 等人的研究結果雖然顯示「表徵文字系列的知識層級」、「計算次數」以及「關係複雜度」等三個變項足以預測系列完成作業的難度，但這三者之間可能存在的線性相依現象卻是該篇研究未曾處理的問題。

從上述討論可知，目前對於系列完成作業的研究尚有待改進之處，尤其是各篇研究中所處理的研究變項對於試題難度的影響也須由統整的觀點來驗證。近來 LLTM 的發展已有部分研究者改以可觀察的試題成分爲分析主體，因此藉由此模式來了解影響系列完成作業試題難度的試題成分應爲可行。

研究方法

一、研究樣本

本研究樣本資料係取自青年智力測驗中的數學測驗預試樣本；該樣本取自大台北地區十所高中二年級學生，有效樣本共計 446 名（郭生玉、林世華、蘇宜芬、宋德忠，民87）。

二、研究工具：數學測驗

(一)編製依據

本測驗係參考 Butterfield 等人（1985）、Holzman 等人（1983），以及 LeFevre 與 Bisanz（1986）的研究設計而成，用以了解個體發現數字系列間所隱藏的關係、並據此解題的推理能力。本測驗共計 50 題，以二元計分方式計分，滿分爲 50 分。試題以選擇的方式呈現，其中 38 題要求受試者從給定的數列推論其間可能隱藏的關係，並且利用這個關係，從選項中選擇正確的數字，使所發現的關係得以繼續發展；另外 12 題乃經刻意破壞之不規則數列，無法形成某一特定的規則，因此受試者須判斷其爲「無法形成規則」才能得分。本測驗編製步驟詳述如下。

(二)編製步驟

1.決定字串數目：根據 Simon 與 Kotovsky（Simon & Kotovsky, 1963; Kotosky & Simon, 1973），Butterfield 等人（1985）以及 Holzman 等人（1983）的研究設計可知，數字系列作業與文字系列作業是由更小的單位---字串---所組成，而本測驗題目則皆由單一字串

以及兩個字串兩類題型組成，當數列是由單一字串所組成時，其週期（period）即由該字串決定；當數列是由兩個字串所組成時，其週期即由兩個字串共同決定。

2.決定字串隱含之關係：Holzman 等人（1983）的研究指出，數字系列中所涉及之運算種類以及數值大小對於受試者答題表現具有相當的影響。爲避免受試者數學先備能力影響其答題表現，本測驗編製時乃參考 Butterfield 等人（1985）對文字系列的研究，限定各字串數字隱含之關係由 N（+1）、I（相同）、B（-1）三者組成。本測驗題目可簡單分類爲單一字串數列以及與兩個字串數列，在單一字串數列中其隱含的數字關係設定由電腦隨機抽選 3 個至 5 個 N、I、B 關係組成，例如：NBN、NIBB、NIBBI。在兩個字串數列中，爲考慮版面限制，以及避免受試者因試題週期過長而無法在三個循環內尋找出其隱含的關係，因此決定分別採用 2 個與 3 個 N、I、B 關係組成，例如規則 N1N2I1N2B2 即由兩個字串所組成，其中字串 1 所隱含的關係爲 NI，字串 2 所隱含的關係爲 NNB。

3.決定數列規則：自先前步驟可知無論是單一字串數列或是兩個字串數列其隱含關係皆是由 3 個至 5 個 N、I、B 所組成。在抽取出 3 個至 5 個 N、I、B 關係後，即由電腦將這些關係加以隨機排列組成各種不同的字串，至此各個數列的規則也皆已確定。

4.決定各字串之起始數字：本測驗題目經由先前的步驟已大致成形，再給予各個規則一個適當的起始數字便可產生無數具有相同規則的數列。爲避免數值大小影響受試者解題表現，本測驗設定由電腦隨機抽取 18 至 42 之間的數字作爲試題的起始數字，因而全體試題之數值大小皆介於 10 與 50 之間。

5.產生配對題：爲提供試題分析另一參考指標，測驗編製時每一規則皆給予兩個不同的起始數值，因此在原始題目中每個規則都有兩題配對題以供參考比較。

6.刪除不合宜試題與決定題目之相對位置：綜合 Butterfield 等人（1985），Holzman 等人（1983），Simon 與 Kotovsky（Simon & Kotovsky, 1963, Kotovsky & Simon, 1973）以及 LeFevre 與 Bisanz（1986）等研究者對於系列完成作業的探討可知，解決有關記憶數列（例如：34 35 36 37 38 39 …、40 40 41 41 42 42 …）以及非移動性數列（22 23 24 22 23 24 …、35 34 33 34 35 34 33 34 …）問題時僅需以視覺判斷，或是與長期記憶中的記憶數列對照即可作答。由於本測驗各試題隱含之規則與起始數字皆是由電腦隨機產生，初步產生之試題包含部分記憶數列以及非移動數列等較不合宜試題，因此測驗設計者以人工判斷之方式篩檢適當的題目，至此共產生 50 道題目。此外，由於在解決兩個字串系列作業相較於單一字串系列作業需要更多工作記憶量，爲使受試者能較快熟悉測驗題目，測驗設計者設定前 24 題爲單一字串數列，後 26 題爲兩個字串數列。而爲避免相同規則試題過於鄰近而產生學習效應，不論單字串題或兩個字串題之試題位置皆由電腦隨機分派。

7.產生不規則數列：由於本測驗以選擇的方式呈現，每題的選目有四個，若平均分

配選目出現的次數，每個選目出現的次數應為 12 至 13 次，而測驗設計者於編製之初已設定每個題目的第四個選目固定為「無法形成規則」，因此整體試題中應有 12 至 13 題為不規則數列。為產生不規則數列，測驗設計者自先前 50 道試題中選取 12 道題以人工的方式進行數列的破壞。所謂數列的破壞乃自正確數列中隨機選取 1 個數字，將其數值大小以「+1」或「 -1」的原則做小幅度的破壞（例如：正確數列的數字為 43，則以 42 或是 44 取代以破壞試題規則）。經過破壞的 12 道試題中有 6 道題為單一字串數列，6 道題為兩個字串數列。

8.決定答案欄所在位置與設計選目：為避免受試者因數列循環週期過少無法尋找出其隱含的關係，試題設計者安排每一試題答案欄於該試題的第三個至第五個循環週期之中。在挑選完答案欄所在位置後，隨即進行選目的設計。為使選目具有誘答性，每個錯誤選目的數值皆配合正確答案所在位置，使其符合數字邏輯。

從上述數學測驗編製步驟可知，本測驗的試題成分除了字串個數、字串所隱含的關係、以及規則破壞與否等三個試題成分為試題設計之初已先行確定外，其餘試題成分皆由電腦隨機產生。研究者參考 Butterfield 等人（1985）、Holzman 等人（1983）、LeFevre 與 Bisanz(1986)以及 Simon 與 Kotovsky(Simon & Kotovsky, 1963, Kotovsky & Simon, 1973）等人對系列完成作業試題特性的研究，逐題分析數學測驗，共計得到八種試題成分：「字串個數」、「週期長度」、「是否為規則數列」、「是否為移動數列」、「相同數字是否鄰近」、「單一字串中是否有相同數字」、「單一字串中相同數字個數與該字串週期長度是否相等」、「字串間是否有假造相同的數字」。這八種試題成分中，「週期長度」、「是否為移動數列」、「相同數字是否鄰近」、「單一字串中相同數字個數與該字串週期長度是否相等」、「字串間是否有假造相同的數字」等五種成分對於數字系列完成作業試題難度的影響已陸續在過去的研究被討論，本研究除了繼續討論這些試題成分對於試題難度的影響外，並討論「字串個數」、「單一字串中是否有相同數字」與「規則或不規則數列」等三種成分對於試題難度的影響。

本測驗之編製雖經由電腦隨機產生題目來減少人為主觀操弄的因素，但這樣的方式卻也造成 10 個試題週期長度無法確定，以及全體試題中單一字串裡相同數字個數與該字串週期長度皆不相等。由於本研究目的之一在了解試題成分對於試題難度的影響，因此研究者僅針對數學測驗中 40 個週期長度確定的題目作分析，但因為此 40 題在「單一字串中相同數字個數與該字串週期長度是否相等」一試題特徵上並不具變異性，因此本研究將不分析此試題成分對於試題難度的影響。有關此 40 題試題成分之統計請見表一。

三、資料分析

㈠在試題反應與模式適合度考驗方面：

1.為了解數學測驗之試題反應資料是否呈單一維度，以 TESTFACT（Wilson, Wood, & Gibbons, 1983）電腦程式中完全訊息（full information）因素分析方法分析測驗反應資料之維度。

2.為了解數學測驗之試題反應資料是否符合 Rasch 模式，以 BILOG 3 電腦程式進行適合度考驗。

㈡在試題難度成分分析方面：

1.為了解數學測驗試題成分對試題難度的預測迴歸公式，以 SPSS for Windows 電腦套裝軟體進行多元逐步迴歸分析。

2.為了解數學測驗之試題反應資料在 LLTM 上的表現，以 LINLOG（Embretson, 1989）電腦程式進行試題分析。

以上各項統計考驗中，除了適合度考驗以 α=.001 為顯著水準進行假設考驗外，其餘各項統計分析皆以 α=.05 為顯著水準進行假設考驗。

表一　數學測驗正式分析試題之試題成分分析統計

字串個數（S）		週期長度（P）			
S=1 24 題	S=2 16 題	P=3 8 題	P=4 16 題	P=5 8 題	P=6 8 題
是否為規則數列		是否為移動數列			
是 31 題	否 9 題	是 34 題		否 6 題	
數列相同數字是否鄰近		單一字串中是否有相同數字			
是 24 題	否 16 題	是 36 題		否 4 題	
單一字串中相同數字個數與該字串週期長度是否相等		不同字串出現假造相同的數字			
是 0 題	否 40 題	是 8 題		否 32 題	

研究結果與討論

一、因素分析結果與討論

數學測驗完全訊息因素分析結果以及各個試題的因素負荷量如表 2 與 3 所示。有關因素分析結果的解釋以及因素個數的決定至今尚未達成共識，Zwick（1987, p.303）表示

在利用 TESTFACT 軟體進行完全訊息因素分析後，若過於依賴卡方檢定所提供的指標，則容易發生因素個數過多的情形，故此因素個數的決定仍須依賴於傳統心理計量所使用的指標，諸如因素負荷量、各因素所能提供的變異解釋量、特徵值的大小等來作爲取決的依據。Reckcase（1979）表示因素分析後，第一個因素的特徵值顯著大於第二個因素的特徵值，且第一個因素所能解釋的變異量大於 20%時，則該資料大致呈單維度；Lord 認爲第一個特徵值比第二個特徵值大的很多，且第二個特徵值比其他的特徵值大的不多時，測驗大致成爲單維度（引自王寶墉，民 84，p.38）。從表二可知，不論是在投入單一因素、兩個因素、或是五個因素時，因素 1 所能提供的解釋量皆達 27%以上，第一個特徵值約爲第二個特徵值與第三個特徵值的三倍以上，並且第二個特徵值與第三個特徵值兩者間相距也不大，因此本資料符合 Reckcase（1979）與 Lord 兩人所提之標準，易言之，本測驗反應資料應可視爲單一維度。

表二　數學測驗之完全訊息因素分析摘要表（N=446）

因素個數	前五大特徵值					個別因素所提供之變異百分比					整體卡方值	卡方改變值
1	11.95	4.00	2.50	0.94	0.67	27.12					8374.29***	
2	12.08	4.20	2.58	1.00	0.79	27.02	8.98				8081.07***	293.22***
3	12.13	4.26	2.73	1.08	0.83	28.24	8.36	7.17			7955.37***	125.70***
4	12.14	4.27	2.77	1.19	0.86	27.83	9.17	8.08	2.71		7756.25***	199.12***
5	12.17	4.30	2.79	1.22	0.94	29.05	10.69	7.39	3.07	2.49	7706.67***	49.58

*** $p<.001$

表三各個試題的因素負荷量顯示，除了第 14 題在本測驗上的因素負荷量較低外，其餘試題因素負荷量皆在 .2 以上。再深入比較規則數列與不規則數列兩類試題因素負荷量的差異時，可發現 9 題不規則數列與 31 題規則數列在因素負荷量的表現上，若排除第 14 題，則兩類試題並未有顯著的區別，並且 9 題不規則數列並未明顯形成另一因素。綜合表二與表三的結果可得以下結果：不論是規則數列或是不規則數列，其所測量之心理構念皆可視爲相同維度；易言之，本測驗試題反應資料大致呈現單一維度，符合 IRT「單一維度」之基本假設，惟獨造成第 14 題因素負荷量偏低的原因仍待進一步的討論。

由表三可知第 14 題的因素負荷量明顯偏低（.078），似乎意指第 14 題所測量之心理構念與其他試題有所不同。研究者在檢視其配對題（即第 20 題）於數學測驗上因素負

荷量的表現時，發現該配對題在因素負荷量上的表現頗佳（.600），而兩者之間唯一的差別在於第 14 題屬於不規則數列，第 20 題屬於規則數列，因此有可能是規則的破壞造成第 14 題在因素負荷量上的表現不佳。再回頭檢視整體試題時，研究者發現在第 14 題之前有 4 題不規則數列（第 4、5、12、13 題），這四題出現了下述雷同的表現：第 4 題與第 5 題爲連續兩題不規則數列題，而第 12 題與第 13 題也是連續兩題不規則數列題，其共同的表現是第 5 題的因素負荷量比第 4 題低，而第 13 題的因素負荷量比第 12 題低；易言之，在連續性的不規則數列出現時，越是後面的題目因素負荷量越有可能偏低。而第 14 題正好是續接著第 12 題與第 13 題，在連續三題爲不規則數列時，最後一題（即第 14 題）的因素負荷量表現正好最低。

表三　數學測驗之試題完全訊息因素負荷量一覽表

題號（配對題）	因素負荷量	題號（配對題）	因素負荷量
1（7）	0.421	21（19）	0.438
2（18）	0.324	22（10）	@0.572
3（8）	0.537	23（6）	0.390
4（24）	@0.357	24（4）	0.701
5（15）	@0.264	25（49）	0.561
6（23）	0.495	26（50）	0.446
7（1）	0.448	28（36）	0.659
8（3）	0.493	29（45）	@0.416
9（17）	0.753	31（41）	0.615
10（22）	0.693	32（42）	0.678
11（13）	0.248	33（47）	0.619
12（16）	@0.453	35（38）	@0.729
13（11）	@0.390	36（28）	0.621
14（20）	@0.078	38（35）	0.710
15（5）	0.473	41（31）	0.630
16（12）	0.425	42（32）	0.777
17（9）	0.747	45（29）	0.674
18（2）	0.486	47（33）	@0.277
19（21）	0.482	49（25）	0.574
20（14）	0.600	50（26）	0.667

註：@ 表示不規則數列

表四 數學測驗之 Rasch 模式試題校準（N=446）

題號（配對題）	b參數	χ^2值	自由度	題號（配對題）	b參數	χ^2值	自由度
1（7）	-2.143	3.8	6	21（19）	-1.845	6.0	6
2（18）	-1.721	20.1	7	22（10）	-3.437	3.0	2
3（8）	-2.187	3.1	6	23（6）	-1.977	9.5	6
4（24）	-3.091	5.7	3	24（4）	-3.276	7.8	2
5（15）	-2.164	10.2	6	25（49）	-1.590	8.3	5
6（23）	-2.143	17.9	6	26（50）	0.656	14.0	7
7（1）	-2.232	6.6	5	28（36）	-1.145	25.4***	6
8（3）	-1.938	9.0	6	29（45）	-2.165	2.0	6
9（17）	-3.843	5.3	1	31（41）	-0.268	21.6	8
10（22）	-3.557	6.5	1	32（42）	-2.121	15.7	4
11（13）	-0.183	32.2***	8	33（47）	-1.364	23.9***	6
12（16）	-2.482	8.0	6	35（38）	-2.483	18.4***	3
13（11）	-2.482	7.3	6	36（28）	-1.185	22.1	6
14（20）	-0.439	60.0***	8	38（35）	-2.377	24.7***	3
15（5）	-1.688	10.9	7	41（31）	-0.247	26.5***	8
16（12）	-1.145	14.5	7	42（32）	-1.791	45.6***	4
17（9）	-4.018	0.0	0	45（29）	-1.068	25.8***	6
18（2）	-1.574	8.1	7	47（33）	-0.428	26.1	8
19（21）	-1.900	14.5	6	49（25）	-1.845	27.5***	4
20（14）	-2.429	3.5	5	50（26）	-0.627	41.5***	6
		平均數			-1.849		
		標準差			1.042		
		總和				642.6***	213

*** $p<.001$

造成上述情形的原因可能在於受試者作答時不敢確定是否有可能連續兩題都是「不規則數列」，因此在遇到此類題目時便對於所謂的「正確答案」產生懷疑態度，這種懷疑正確答案的心態尤其是在面對諸如第 12、13、14 題等連續三題不規則數列時更有可能出現。因此，第 14 題因素負荷量的表現不佳，有可能是因爲其爲連續三題不規則數列中

的最後一題所造成，而並非單純因爲其爲不規則數列。

表五　數學測驗之二參數模式試題校準（N=446）

題號（配對題）	a參數	B參數	χ^2值	自由度	題號（配對題）	a參數	b參數	χ^2值	自由度
1（7）	0.510	-2.483	2.3	5	21（19）	0.464	-2.307	9.8	7
2（18）	0.418	-2.349	5.6	6	22（10）	0.638	-3.334	2.6	2
3（8）	0.605	-2.224	6.4	5	23（6）	0.429	-2.639	8.5	5
4（24）	0.418	-4.220	3.4	3	24（4）	0.819	-2.679	2.3	1
5（15）	0.374	-3.243	3.7	6	25（49）	0.702	-1.470	9.9	6
6（23）	0.553	-2.330	5.4	5	26（50）	0.730	0.552	10.5	6
7（1）	0.491	-2.665	8.2	6	28（36）	1.060	-0.849	6.2	6
8（3）	0.543	-2.142	4.5	5	29（45）	0.485	-2.611	6.0	6
9（17）	1.055	-2.690	0.5	1	31（41）	1.096	-0.218	3.0	5
10（22）	0.858	-2.815	2.0	1	32（42）	1.337	-1.384	1.0	2
11（13）	0.281	-0.400	5.3	9	33（47）	1.145	-0.969	4.3	5
12（16）	0.476	-3.035	9.8	5	35（38）	1.208	-1.671	3.4	2
13（11）	0.422	-3.358	7.3	5	36（28）	1.093	-0.865	5.9	6
14（20）	0.181	-1.327	8.1	9	38（35）	1.385	-1.520	1.3	2
15（5）	0.465	-2.111	9.3	7	41（31）	1.171	-0.198	7.6	5
16（12）	0.489	-1.387	8.6	7	42（32）	2.113	-1.038	1.4	2
17（9）	1.135	-2.702	0.0	0	45（29）	1.349	-0.718	0.6	5
18（2）	0.467	-1.965	5.9	7	47（33）	0.320	-0.771	12	8
19（21）	0.517	-2.181	14.3	6	49（25）	1.207	-1.264	7.2	4
20（14）	0.639	-2.368	6.8	5	50（26）	1.471	-0.424	6.4	4
			平均數			0.778	-1.859		
			標準差			0.419	1.045		
			總和					227.4	192

在 LeFevre 與 Bisanz（1986）的研究中，安排不規則數列的目的在於了解數字系列解題歷程中的「關係偵測」階段是否尚隱含著其他認知歷程。該篇研究雖然提供測驗設

計另一思考方向，但卻也同時引發另一問題：規則破壞與否是否會影響試題的維度？LeFevre 與 Bisanz（1986）的研究顯然未回答這個問題，但從本研究資料分析結果可知：偵測數列規則數列與不規則數列的心理能力是同一維度的，因此在編製數字系列完成測驗時，加入不規則數列試題是可接受的，惟獨當不規則數列試題若發生連續出現時，有可能出現因素負荷量下降的情形，此爲測驗編製時需注意的問題。

二、潛在特質模式適合度分析結果與討論

表四與表五分別爲數學測驗在 Rasch 模式以及二參數模式的試題校準。表四顯示本測驗中第 11、14、28、33、35、38、41、42、45、49、50 等 11 個題目在 Rasch 模式上的適合度未盡理想（卡方值達.001 顯著水準），初步探討其原因時，發現這些題目皆適配於二參數模式（卡方值未達.001 顯著水準），亦即其鑑別度有所差異。

在深入探討此 11 題未適配題與另外 29 題適配題在的試題成分後，研究者發現這兩者之間在試題組成成分上並未出現明顯的差異，因此造成未適配試題鑑別度差異應可排除試題成分之因。再度檢視這 11 個題目與其配對題上的差異時，則發現這些配對題大多適配於 Rasch 模式，由此看來以相同規則建造出的兩個試題分別適配於不同模式，頗能確定這背後的原因並非試題成分所造成。

三、試題難度成分分析

(一)多元逐步迴歸分析結果與討論

表六與表七爲數學測驗試題成分對試題難度的多元逐步迴歸分析結果（以下簡稱爲 LR 分析，Linear Regression Analysis），其中預測變項爲試題成分（參見表 10），效標變項爲 Rasch 模式之 b 參數。由於本測驗中每個試題在試題成分「單一字串中相同數字個數與該字串週期長度是否相等」上的表現皆爲不相等，不具有任何變異性，因此在 LR 分析時並不投入此試題成分。由表六可知「是否爲移動數列」（X4）、「週期長度」（X5）、「字串間是否有假造相同的數字」（X6）、「字串個數」（X7）等四類試題成分與 b 參數具有顯著相關，但是在 LR 分析時只有「是否爲移動數列」（X4）、「週期長度」（X5）與「字串間是否有假造相同的數字」（X6）三類試題成分對於 b 參數具有正向預測力：其多元相關係數爲.822，並能提供 64.8%的變異解釋量，而其原始分數迴歸方程式與標準化分數迴歸方程式分別爲

$$\hat{Y} = 0.702X_4 + 0.551X_5 + 0.821X_6 - 5.033 \quad \text{公式（1）}$$

$$\hat{Z}_Y = 0.244Z_4 + 0.546Z_5 + 0.319Z_6 \quad \text{公式（2）}$$

表六　數學測驗試題成分、試題難度相關係數與標準化迴歸係數（β）一覽表

	B_1PL	X1	X2	X3	X4	X5	X6	β
X1	-.093							-.068
X2	-.147	-.049						-.032
X3	.073	-.408**	-.020					.180
X4	.476***	-.343*	-.109	-.140				.244*
X5	.713***	-.180	-.094	-.196	.302*			.546***
X6	.531***	.408**	-.120	-.167	.210	.294*		.319**
X7	.473***	-.167	-.073	-.408**	.343*	.480***	.612***	-.136

註　B_1PL：Rasch 模式 b 參數； X1：相同數字是否鄰近； X2：是否爲規則數列；X3：單一字串中是否有相同數字；X4：是否爲移動數列； X5：週期長度； X6：字串間是否有假造相同的數字； X7：字串個數； *P<.05； **P<.01； ***P<.001

表七　數學測驗多元逐步迴歸分析變異數分析摘要表（N=446）

來源	SS	df	MS	F	R	R^2
迴歸	28.570	3	9.523	24.910***	.822	.675
殘差	13.763	36	.382			
全體	42.334	39				

註　預測變項：是否爲移動數列（X4）、週期長度（X6）、字串間是否有假造相同的數字（X7）；效標變項：Rasch 模式 b 參數； *** p<.001

公式（1）與公式（2）顯示「是否爲移動數列」（X4）、「週期長度」（X5）與「字串間是否有假造相同的數字」（X6）三類試題成分對於 b 參數具有正向的影響力：當試題具有三種試題成分之一時，其 b 參數估計値將上升；易言之，題目之難度將提高。這樣的研究結果與過去的研究互有一致與不一致之處，以下將針對本研究所探討的七個試題成分分析其與試題難度間的關係。

1.就試題成分「是否爲移動數列」對於試題難度的影響而言

Simon 與 Kotovsky（Simon & Kotovsky, 1963；Kotovsky & Simon, 1973）的研究中曾指出，關係 I 之偵測比關係 N 與關係 B 容易；Holzman 等人（1983）則主張非移動數列

的工作記憶需求量（working memory demand）爲零，因此其最容易爲受試者解決；Butterfield 等人（1985）甚且認爲在處理非移動系列時僅需以視覺判斷便可以作答。在本研究結果中，「是否爲移動數列」一試題成分與試題難度的相關爲.476 （p<.001），並且對於試題難度具有預測力，其標準化迴歸係數爲.244（p<.05），顯示當試題爲移動數列時，其難度較高；當試題爲非移動數列時，其難度較低，如此的結果與 Simon 與 Kotovsky、Holzman 等人以及 Butterfield 等人的看法可說是一致。

2.就試題成分「週期長度」對於試題難度的影響而言

「週期長度」爲系列完成作業中最顯而易見的試題成分，而部分研究者（Simon & Kotovsky, 1963；Kotovsky & Simon, 1973；Holzman *et al.*, 1983）認爲兩者之間應具有正相關的關係，但過去的研究並未全然支持此論點（Simon & Kotovsky, 1963; Kotovsky & Simon, 1973; Holzman *et al.*, 1983; Butterfield *et al.*, 1985）。本研究資料分析結果顯示「週期長度」與試題難度的相關爲 .713 （p<.001），這表示週期長度越長，則試題難度越高；此外，本研究結果並顯示週期長度對於試題難度具有預測力，其標準化迴歸係數爲.610。這樣的研究結果雖然與 Simon 與 Kotovsky、Holzman 等人，以及 Butterfield 等人研究結果不符，但卻與 Simon 與 Kotovsky 以及 Holzman 等人的看法一致。

在比較本研究與 Simon 與 Kotovsky、Holzman 等人，以及 Butterfield 等人研究設計上的差異後，研究者發現以往研究所分析的系列問題週期長度侷限於 2~4 之間，而本研究所分析之數學測驗的週期長度共計有 3、4、5、6 等四種。在解決系列問題時，規則週期的察覺是需要足夠的循環次數才有可能。以數字系列完成作業爲例，在版面的限制之下，若一頁的寬度能出現 20 個數字，對於週期爲 4 的題目而言，其可能出現的循環次數將近爲 5 次，但是對於週期爲 6 的題目而言，其可能出現的循循次數卻只有 3 次，對於具有相同規則的數列，循環次數越多，其關係越容易被偵測出乃是必然的現象，反之，則受試者在有限的線索之下做出正確反應的機率也將更小。

此外，以 Rasch 模式 b 參數爲依變項，「週期長度」與「相同數字是否鄰近」爲自變項進行二因子變異數分析結果顯示「週期長度」與「相同數字是否鄰近」兩者間存在交互作用（F=15.026, p<.001），這表示當試題週期越長，且相同的數字不鄰近時，其試題難度將提高。

綜合上述討論可知，「週期長度」對於試題難度的影響，除了其本身的特性使然外，該成分與「相同數字是否鄰近」也會對試題難度產生交互作用的影響。

3.就試題成分「字串間是否有假造相同的數字」對於試題難度的影響而言

試題成分「字串間是否有假造相同的數字」最先爲 Simon 與 Kotovsky（Simon & Kotovsky, 1963; Kotovsky & Simon, 1973）所注意，在其研究中顯示關係 I 雖然比關係 B 與關係 N 容易爲受試者所偵測出，但是當系列包含著假造相同的字母或數字時，關係 I

的存在反而會增加試題的難度。本研究分析結果顯示「字串間是否有假造相同的數字」一試題成分與試題難度相關為 .531（$p<.001$），並且對於試題難度具有正向預測力，其標準化迴歸係數為 .319（$p<.01$）。如此的研究結果不但與 Simon 與 Kotovsky（Simon & Kotovsky, 1963; Kotovsky & Simon, 1973）所提之論點一致，並且也與 Butterfield 等人（1985）的研究結果相符。在 Butterfield 等人（1985）的研究也顯示，字串間若出現假造相同的字母時，其試題難度將相度增加，並且此試題成分對於試題難度具有正向預測力。

研究者認為「字串間是否有假造相同的數字」之所以會促使試題難度增加是因為當字串間存在假造相同的數字時，將使得整個系列以及個別字串的週期辨識更加困難。Simon 與 Kotovsky 的研究中便曾提到，一旦受試者無法順利尋找出系列的週期，則答對該系列的可能性幾近為零。以數列 S_1 "20 20 19 21 21 18 22 22 17...."以及數列 S_2"20 20 20 21 21 20 22 22 20 23 23 20...."為例，數列 S_1 中相同的數字 20、21、22 在整個系列間重複出現，而數列 S_2 中相同數字 20 也重複出現在整個系列之間，但是數列 S_1 中相同的數字 20、21、22 的功能卻是使得週期更容易辨識（鄰近法），而數列 S_2 中假造相同的數字 20 卻是使得數列的週期變得更模糊不易辨識。因此研究者認為「字串間是否有假造相同的數字」一試題特徵之所以對於試題難度具有決定性的影響是主要是因為該試題特徵會影響到週期的辨識。

4.就試題成分「字串個數」對於試題難度的影響而言

「字串個數」一試題成分首先為 Butterfield 等人（1985）所提出，不過 Butterfield 等人提出此試題成分之目的在於解說試題規則的構成，而不在用以說明試題難度的來源。本研究分析結果顯示，字串個數與試題難度兩者間具有正相關（$r= .473$, $p< .001$）；易言之，字串數越多，則試題難度越高。但在深入了解字串個數與試題難度的關係後，研究者發現這兩者之間的關係尚牽涉到其他的因素。以本研究所分析的數學測驗而言，前 24 題為單一字串題，後 16 題為兩個字串題。就單字串題而言，其週期長度為 3、4、5，而兩個字串題的週期長度為 4 與 6。從先前的討論可知，週期長度對於試題的難度具有正相關，並有相當的預測力，週期長度分配上的不等有可能是促使兩個字串題較單字串困難的原因。

此外，本測驗中另外一個具有預測力的試題成分「字串間是否有假造相同的數字」也可能是促使兩個字串題較單字串題困難之原因。所謂「字串間是否有假造相同的數字」是指在不同字串中出現相同的數字，因此其前提必須有兩個（或以上）的字串才有可能出現假造相同的數字。本測驗的類別侷限於單字串數列與兩個字串數列，在單字串數列中是不可能出現假造相同數字的情形，而兩個字串數列中出現假造相同數字的題目多達 8 題，這兩者之間的差別也可能造成兩個字串數列較單字串數列困難。

除了上述因素外，數學測驗中兩個字串題皆為移動數列，而「是否為移動數列」除

了與難度具有正相關外，並能預測試題難度，有可能藉由如此間接的歷程，而使得「字串個數」與試題難度兩者之間具有顯著的相關。綜合上述之推論可知，在本研究中「字串個數」與試題難度具有正相關的可能因素是多重的，雖然 Butterfield 等人的研究並未支持「字串個數」與試題難度兩者可能存在的關係，但是就本研究而言，這樣的結果應屬合理。

5.就試題成分「相同數字是否鄰近」對於試題難度的影響而言

Holzman 等人（1983）的研究中曾提及，用以發現數列週期的方法有「鄰近法」（adjacent approach）與「非鄰近法」（nonadjacent approach）兩種，其中鄰近法較非鄰近法易於發現週期，易言之，當相同的數字正好鄰近時，則週期長度較易於確認。Butterfield 等人（1985）也主張，當相同的文字比鄰時，受試者易於將其視爲同樣的區組，因此試題難度將降低。這兩位學者的皆傾向認爲相同數字鄰近時試題難度會下降，然而本研究分析結果卻顯示在數學測驗中「相同數字是否鄰近」與試題難度的相關並不明顯（$r=-.093, p>.05$）。

研究者進一步利用 Rasch 模式 b 參數爲依變項，「相同數字是否鄰近」與「週期長度」爲自變項進行二因子變異數分析後，發現「相同數字是否鄰近」與「週期長度」兩者之間具有交互作用（$F=6.675, p<.001$），因此「相同數字是否鄰近」對於難度的影響須視「週期長度」而定（詳見表八）。此外，「相同數字是否鄰近」與「字串個數」兩者之間亦具有交互作用（$F=10.338, p<.01$），因此「相同數字是否鄰近」對於難度的影響亦須視「字串個數」而定。

表八 「相同數字是否鄰近」試題與「不同週期長度」試題之難度平均數摘要表

試題成分	週期長度=3	週期長度=4	週期長度=5	週期長度=6	合
相同數字不鄰近	-2.1 （N=2）	-2.1 （N=6）	-1.5 （N=4）	-1.3 （N=4）	-1.7 （N=16）
相同數字鄰近	-3.5 （N=6）	-1.7 （N=10）	-1.9 （N=4）	-1.2 （N=4）	-1.9 （N=24）
合	-3.2 （N=8）	-1.8 （N=16）	-1.7 （N=8）	-0.7 （N=8）	-1.8 （N=40）

上述兩個二因子變異數分析結果顯示，「相同數字是否鄰近」與試題難度的關係尙須視「週期長度」以及「字串個數」兩試題成分的影響而定。若相同的數字是出現在同一字串，則該字串的週期長度會因爲這些相同數字的存在而更容易辨識（鄰近法），但如果相同的數字是發生在不同字串，則不同字串中相同的數字（亦即「假造相同的關係」）反而會提高試題難度。同樣以數列 S_1“20 20 19 21 21 18 22 22 17....”以及數列 S_2“20 20

20 21 21 20 22 22 20 23 23 20....”爲例，數列 S_1 中相同的數字是出現在同一字串，因此利用鄰近法便能很快地先找出該字串的週期，更進而找出整個系列的週期；而數列 S_2 中假造相同的數字 20 卻是使得數列的週期變得更模糊不易辨識。也因此，單視「相同數字是否鄰近」是難以決定試題難度的，因爲該試題成分尚且會與「週期長度」以及「字串個數」發生交互作用。

表九 「相同數字是否鄰近」試題與「不同字串個數」試題之難度平均數摘要表

試題成分	字串個數=1	字串個數=2	合
相同數字不鄰近	-1.7 （N=8）	-1.7 （N=8）	-1.7 （N=16）
相同數字鄰近	-2.5 （N=16）	-0.7 （N=8）	-1.9 （N=24）
合	-2.2 （N=24）	-1.3 （N=16）	-1.8 （N=40）

6.就試題成分「是否爲規則數列」對於試題難度的影響而言

LeFevre 與 Bisanz（1986）的研究中曾比較規則數列、無效數列（invalid number series）與不規則數列（anomalous number series）三者之間反應時間的差異，其中所謂的無效數列係將規則數列予以小幅度的破壞，不規則數列則是將規則數列予以大幅度的破壞。該篇研究指出，就計數數列而言，無效數列反應時間較規則數列長，而規則數列反應時間又較不規則數列長。若就非計數數列而言，無效數列所需反應時間最長，規則數列次之，不規則數列所需反應時間最短。由於本研究所分析之數學測驗中的不規則數列乃是將規則數列予以＋1 或－1 的小幅度破壞，因此按照 LeFevre 與 Bisanz（1986）的分類應屬於無效數列。

數學測驗的分析結果顯示「是否爲規則數列」一試題特徵與試題難度之相關爲-.147（ $p>.05$ ），這樣的數據雖未達統計上的顯著水準，但卻暗示著規則的破壞可能會促使試題難度的降低。如此的研究結果似乎是與 LeFevre 與 Bisanz（1986）的研究結果相違背的，但是在深入比較兩個研究設計上的差異便能了解其中的原因。爲避免受試者數學先備能力干擾其在數學測驗上的表現，在本研究中使用的數列題材所涉及的關係侷限於 I、N、B 三者，在這樣的數字關係組合下，數學測驗中的字串全爲「記憶字串」（即計數字串）。一旦受試者能夠尋找出該數列所組成的字串，則運用「記憶字串的再認」、「檢視」與「自我終止」策略便能很快判斷出試題的正確答案。以系列 S_1“32 33 34 33 34 35 34 35 36 35 35 37....”爲例，一旦受試者能發覺其乃由週期長度=3 的單一字串所組成的數

列，則受試者只要依此規則順數便能很快地辨識出最後一個週其中有不規則的數字包含在其中。

LeFevre 與 Bisanz（1986）的研究雖然顯示這樣的認知歷程是出現在不規則數列（大幅度規則破壞），但是在本研究中所使用多為記憶字串的情況之下，規則的破壞可能反而促使受試者運用「記憶數列的再認」、「檢視」與「自我終止策略」等認知策略而使得題目變簡單。因此，本研究結果雖然與 LeFevre 與 Bisanz（1986）的研究結果不符，但這兩者間的差異在於試題的型態，而非概念上的差異使然。

7.就試題成分「單一字串中是否有相同數字」對於試題難度的影響而言

試題成分「單一字串中是否有相同數字」的類似概念對於試題難度的影響曾先後於 Simon 與 Kotovsky（Simon & Kotovsky, 1963; Kotovsky & Simon, 1973），Holzman 等人（1983），Butterfield 等人（1985）的研究被討論。在 Simon 與 Kotovsky（Simon & Kotovsky, 1963; Kotovsky & Simon, 1973）的研究中曾提及受試者係藉由發現關係 I 的存在而發現週期，並且關係 I 比關係 B 與關係 N 容易為人所覺察，不過在某些含有假造相同字母的系列中，關係 I 的存在反而讓週期的觀察變得更困難；Holzman 等人（1983）主張在觀察數列的週期時，「鄰近法」比「非鄰近法」容易找出數列的週期，其中鄰近法乃利用系列中重複的數字作為線索；Butterfield 等人（1985）的研究指出由相同字母所組成的非移動系列最易為受試者覺察，不過若不同字串間出現假造相同的字母時，試題難度則相對上升。

綜合上述研究結果可知，「單一字串中是否有相同數字」可以說是一個複雜的概念，與其相關的概念至少尚包括「相同數字是否鄰近」以及「不同字串間是否出現假造相同的數字」。以往的研究多未直接探討「單一字串中是否有相同數字」對於試題難度的影響，有可能真正影響試題難度者不在於「單一字串中是否有相同數字」，而是在於單一字串中相同的數字是呈現如何的組合型態。本研究結果顯示「單一字串中是否有相同數字」與試題難度間的相關為 .073 （$p>.05$），似乎意味在數學測驗中，若該數列出現相同的數字，則試題難度有提高的可能性。

研究者認為「單一字串中是否有相同數字」對於試題難度的影響需視單一字串中相同的數字是呈現如何的組合型態。以系列 S_1‘10 11 40 11 12 41 12 13 42....’、系列 S_2‘10 10 40 11 11 41 12 12 42....’，以及系列 S_3‘10 10 14 11 11 13 12 12 12....’等三個系列為例，這三個系列中單一字串皆包含有相同的數字，但是在比較 S_1 與 S_2 時，就可以發覺 S_2 中相同的數字是鄰近的，所以可以馬上判斷出週期與規則。在比較 S_2 與 S_3 時，又可以發覺 S_3 中最後 3 個相同的數字反而使得週期的辨識變得模糊，因為在 S_3 中除了單一字串有相同的數字外，不同字串間更出現假造相同的數字，因此 S_3 又較 S_2 困難。再細究整個數學測驗時，可發現雷同於 S_1、S_2、S_3 這三類試題的組合普遍散佈在整個測驗。因此，在數

學測驗中「單一字串中是否有相同數字」對於試題難度的影響可能尚需視單一字串中相同數字的組合型態才能決定，而無法單由該試題成分決定。

表十　數學測驗試題成分（η）與多元逐步迴歸預測變項（X）一覽表

題號（配對題）	η_1（X1）	η_2（X2）	η_3（X3）	η_4（X4）	η_5（X5）	η_6（X6）	η_7（X7）	題號（配對題）	η_1（X1）	η_2（X2）	η_3（X3）	η_4（X4）	η_5（X5）	η_6（X6）	η_7（X7）
1（7）	1	0	1	1	4	0	1	21（19）	1	0	1	1	5	0	1
2（18）	0	0	1	1	5	0	1	22（10）	1	1	1	1	3	0	1
3（8）	0	0	1	1	3	0	1	23（6）	1	0	1	1	4	0	1
4（24）	1	1	1	0	3	0	1	24（4）	1	0	1	0	3	0	1
5（15）	1	1	1	0	5	0	1	25（49）	0	0	0	1	6	0	2
6（23）	1	0	1	1	4	0	1	26（50）	1	0	1	1	6	1	2
7（1）	1	0	1	1	4	0	1	28（36）	1	0	1	1	4	1	2
8（3）	0	0	1	1	3	0	1	29（45）	1	1	1	1	4	1	2
9（17）	1	0	1	0	3	0	1	31（41）	1	0	1	1	6	1	2
10（22）	1	0	1	1	3	0	1	32（42）	0	0	1	1	4	0	2
11（13）	0	0	1	1	5	0	1	33（47）	0	0	1	1	6	0	2
12（16）	0	1	1	1	4	0	1	35（38）	0	1	0	1	4	0	2
13（11）	0	1	1	1	5	0	1	36（28）	1	0	1	1	4	1	2
14（20）	1	1	1	1	4	0	1	38（35）	0	0	0	1	4	0	2
15（5）	1	0	1	0	5	0	1	41（31）	1	0	1	1	6	1	2
16（12）	0	0	1	1	4	0	1	42（32）	0	0	1	1	4	0	2
17（9）	1	0	1	0	3	0	1	45（29）	1	0	1	1	4	1	2
18（2）	0	0	1	1	5	0	1	47（33）	0	1	1	1	6	0	2
19（21）	1	0	1	1	5	0	1	49（25）	0	0	0	1	6	0	2
20（14）	1	0	1	1	4	0	1	50（26）	1	0	1	1	6	1	2

註　η_1（X1）：相同數字是否鄰近；η_2（X2）：是否為規則數列；η_3（X3）：單一字串中是否有相同數字；η_4（X4）：是否為移動數列；η_5（X5）：週期長度；η_6（X6）：字串間是否有假造相同的數字；　η_7（X7）：字串個數。

㈡ LLTM 分析結果與討論

本研究除了以迴歸分析的方式來了解數字系列中試題成分對於試題難度的影響外，並且採用 LLTM 對於數字系列試題成分做線性分析，以了解不同試題成分對於試題難度的影響力。根據 LLTM 的基本假設，要對試題難度作線性分解，必須先對試題成分有一明確的界定。有關數學測驗之試題成分請見表十。

表十一與表十二皆爲數學測驗在 LLTM 的分析結果；從表 11 可知數學測驗中各個試題成分對於試題難度的加權值，其中「單一字串中是否有相同數字」（η_3）、「是否爲移動數列」（η_4）、「週期長度」（η_5）、「字串間是否有假造相同的數字」（η_6）等四個試題成分對於難度的影響方向是正向的，當試題具有這些成分時，其難度將上升；「相同數字是否鄰近」（η_1）、「是否爲規則數列」（η_2）與「字串個數」（η_7）等三種試題成分則是對難度的影響是負向的，當試題具有這些成分時，其難度將下降。利用這些加權值，可得到題目難度（σ）與試題成分（η）關係式爲：

$$\sigma = -.28688\times\eta_1 - .036756\times\eta_2 + .73496\times\eta_3 + .63617\times\eta_4 + .55714\times\eta_5 + .98217\times\eta_6 - .17868\times\eta_7 - 3.4200 \quad \text{公式（3）}$$

利用上述公式（3），即可計算出表十二中 LLTM 正規化試題難度參數（σ），由於該表中的兩種參數皆已正規化，因此可以直接比較兩者之數值；從該表可知利用 LLTM 所估算出的試題難度參數（σ）與 Rasch 模式正規化試題難度參數（b）兩者之間的相關達.845（$p<.001$）。此外，在 Rasch 模式正規化試題難度參數（b）與 LLTM 試題難度參數（σ）兩者差異比較所得之卡方值爲 756.00 （df=33），顯示利用 LLTM 與利用 Rasch 模式所得之難度參數兩者間是具有顯著差異的。

在此值得注意的是表十二中 Rasch 模式正規化試題難度參數（b）值與先前表四所報導之 Rasch 模式試題校準 b 參數二者之差別。這兩者皆爲數學測驗之難度指標，而表十二所報導之值乃經正規化所得之參數，表四所報導之值則未經正規化，因此兩者意指相同的試題難度，其差異在於兩者建基於不同的量尺上，而非本質上的差異。

此外，表十二有關數學測驗之 LLTM 分析結果亦顯示，在數學測驗中各個試題成分對於試題難度的影響方向並不一致，而 LLTM 模式並未適當匹配 Rasch 模式（$\chi^2=756.00$, df=33）。然而兩模式所推估之難度指數具有高度正相關（r=.845）。由於 LLTM 與本研究所使用的另一分析方法 LR 分析皆在了解試題成分對試題難度的影響，因此比較這兩者之異同將能更清楚地看出在數字測驗中各試題成分對於試題難度影響的相對重要性。

由表十三數學測驗在 LLTM 與 LR 分析中試題成分加權值一覽表可看出一個趨勢：不論是在 LLTM 或是 LR 分析中，「單一字串中是否有相同數字」、「是否爲移動數列」、「週期長度」、「字串間是否有假造相同的數字」等四種試題成分都能使試題難度

表十一　數學測驗在 LLTM 試題成分加權值一覽表

η_1	η_2	η_3	η_4	η_5	η_6	η_7	常數
-.28688	-.036756	.73496	.63617	.55714	.98217	-.17868	-3.4200

註　η_1：相同數字是否鄰近；η_2：是否爲規則數列；η_3：單一字串中是否有相同數字；η_4：是否爲移動數列；η_5：週期長度；η_6：字串間是否有假造相同的數字；η_7：字串個數。

表十二 數學測驗 LLTM 正規化試題難度參數（σ）與 Rasch 模式正規化試題難度參數（b）之比較（N=446）

題號（配對題）	LLTM 難度參數（σ）	Rasch模式 難度參數（b）	題號（配對題）	LLTM 難度參數（σ）	Rasch模式 難度參數（b）
1（7）	-0.28538	-0.31609	21（19）	0.27176	-0.00754
2（18）	0.55864	0.12115	22（10）	-0.87927	-1.60744
3（8）	-0.55563	-0.36145	23（6）	-0.28538	-0.12496
4（24）	-1.51544	-1.26715	24（4）	-1.47869	-1.44943
5（15）	-0.40117	-0.33862	25（49）	0.20214	0.25901
6（23）	-0.28538	-0.31609	26（50）	1.63238	2.47812
7（1）	-0.28538	-0.40802	28（36）	0.51811	0.72585
8（3）	-0.55563	-0.10488	29（45）	0.48136	-0.33862
9（17）	-1.47869	-2.01136	31（41）	1.63238	1.64034
10（22）	-0.84251	-1.72673	32（42）	-0.17717	-0.29384
11（13）	0.55864	1.72823	33（47）	0.93710	0.49625
12（16）	-0.03525	-0.66234	35（38）	-0.94889	-0.66234
13（11）	0.52189	-0.66234	36（28）	0.51811	0.68432
14（20）	-0.32213	1.46203	38（35）	-0.91214	-0.55586
15（5）	-0.36441	0.15647	41（31）	1.63238	1.66237
16（12）	0.00150	0.72585	42（32）	-0.17717	0.04862
17（9）	-1.47869	-2.11528	45（29）	0.51811	0.80706
18（2）	0.55864	0.27563	47（33）	0.90034	1.47331
19（21）	0.27176	-0.06536	49（25）	0.20214	-0.00754
20（14）	-0.28538	-0.60822	50（26）	1.63238	1.26688

R（σ,b）= .845；χ^2 = 756.00　（df=33）

上升，而「相同數字是否鄰近」、「是否爲規則數列」、「字串個數」等三種試題成分則是使試題難度下降，因此這兩種分析方法所得之結果頗爲一致。在 Green 與 Smith（1987）以 The Knox Cube Block Test 爲題材的研究中，也曾比較利用 LR 與 LLTM 兩種方式對試題難度作分解，研究結果也支持利用這兩種方式所得之結果雷同。研究者爲更進一步確認這兩種分析方法之間的關係，利用公式（2）與公式（3）分別求出數學測驗原始分數的試題難度預測值（$\hat{Y}$）（參見表十四）以及 LLTM 正規化試題難度參數（σ）兩者間之相關係數，分析結果顯示 r=.947（p<.01），可見這兩種分析方式所得結果確實具有一致性。

表十三　數學測驗在 LLTM 與 LR 分析中試題成分加權值一覽表

	η_1（X1）	η_2（X2）	η_3（X3）	η_4（X4）	η_5（X5）	η_6（X6）	η_7（X7）	常數
LLTM	-.28688	-.036756	.73496	.63617	.55714	.98217	-.17868	-3.4200
LR（β）	-.068	-.032	.180	.244*	.546***	.319***	-.136	--

註 η_1（X1）：相同數字是否鄰近；η_2（X2）：是否爲規則數列；η_3（X3）：單一字串中是否有相同數字；η_4（X4）：是否爲移動數列；η_5（X5）：週期長度；η_6（X6）：字串間是否有假造相同的數字；η_7（X7）：字串個數；*p<.05；***p<.001。

表十四　數學測驗原始分數試題難度預測值（$\hat{Y}$）一覽表

題號（配對題）	$\hat{Y}$	題號（配對題）	$\hat{Y}$	題號（配對題）	$\hat{Y}$	題號（配對題）	$\hat{Y}$
1（7）	-2.127	11（13）	-1.576	21（19）	-1.576	33（47）	-1.025
2（18）	-1.576	12（16）	-2.127	22（10）	-2.678	35（38）	-2.127
3（8）	-2.678	13（11）	-1.576	23（6）	-2.127	36（28）	-1.306
4（24）	-3.380	14（20）	-2.127	24（4）	-3.380	38（35）	-2.127
5（15）	-2.278	15（5）	-2.278	25（49）	-1.025	41（31）	-0.204
6（23）	-2.127	16（12）	-2.127	26（50）	-0.204	42（32）	-2.127
7（1）	-2.127	17（9）	-3.380	28（36）	-1.306	45（29）	-1.306
8（3）	-2.678	18（2）	-1.576	29（45）	-1.306	47（33）	-1.025
9（17）	-3.380	19（21）	-1.576	31（41）	-0.204	49（25）	-1.025
10（22）	-2.678	20（14）	-2.127	32（42）	-2.127	50（26）	-0.204

結論與建議

根據本研究的結果，歸納結論與建議如次。

一、研究發現與結論

(一)在試題反應與模式適合度考驗方面

1.青年智力測驗中的數學測驗預試反應資料大致呈現單一維度。

2.青年智力測驗中的數學測驗預試反應資料大致符合 Rasch 模式。

(二)在試題難度成分分析方面：

1.青年智力測驗中的數學測驗預試反應資料在 LR 分析中顯示，「是否為移動數列」、「週期長度」、「字串間是否有假造相同的數字」以及「字串個數」等四種試題成分與 Rasch 模式難度參數（b）具有顯著相關，其中「是否為移動數列」、「週期長度」以及「字串間是否有假造相同的數字」三種試題成分對於 Rasch 模式難度參數（b）具有正向預測力，易言之，這三種試題難度將促使試題難度提高。

2.青年智力測驗中的數學測驗預試試題反應資料以 LR 分析所得之試題難度原始分數迴歸方程式與標準化分數迴歸方程式分別為

$$\hat{Y} = 0.702X_4 + 0.551X_5 + 0.821X_6 - 5.033$$

$$\hat{Z}_y = 0.244Z_{x4} + 0.546Z_{x5} + 0.319Z_{x6}$$

上式中 X_4、X_5 與 X_6 分別為「是否為移動數列」、「週期長度」與「字串間是否有假造相同的數字」等試題成分。利用上述公式所得之預測難度與實際難度值之間相關為 .822，顯示預測難度能提供實際難度 64.8%的變異解釋量。

3.青年智力測驗中的數學測驗預試反應資料在 LLTM 與 Rasch 模式兩者間的表現有顯著的差異（χ^2=756.00, df=33），但兩者之間具有高度的正相關（r=.845）。此外，利用 LLTM 所得之試題成分對於試題難度之加權值顯示，「數列中是否有相同數字」、「是否為移動數列」、「週期長度」以及「字串間是否有假造相同的數字」等四種試題成分對於試題難度有正向影響力，亦即這四種試題成分會促使試題難度上升；而「相同數字是否鄰近」、「是否為規則數列」以及「字串個數」等試題成分對於試題難度有負向的影響力，亦即這三種試題成分會促使試題下降。

4.青年智力測驗中的數學測驗預試試題反應資料以 LLTM 分析所得之試題成分對試題難度的關係式為

$$\sigma = -.28688\times\eta_1 - .036756\times\eta_2 + .73496\times\eta_3 + .63617\times\eta_4 + .55714\times\eta_5 + .98217\times\eta_6 - .17868\times\eta_7 - 3.4200$$

上式中η_1至η_7分別爲「相同數字是否鄰近」、「是否爲規則數列」、「數列中是否有相同數字」、「是否爲移動數列」、「週期長度」、「字串間是否有假造相同的數字」、「字串個數」等試題成分。

5.青年智力測驗中的數學測驗預試試題反應資料顯示，就試題成分對於試題難度的影響而言，利用 LR 分析與利用 LLTM 所得之結果具有相當的一致性（$r=.947, p<.01$）。

二、建議

(一)應用本研究結果之建議

1.在測驗編製方面，可藉由操弄試題成分來了解題目難度的變異來源

本研究結果顯示由研究者所提出的「是否爲移動數列」、「週期長度」與「字串間是否有假造相同的數字」等三種試題成分能對數字系列完成作業的試題難度提供相當的變異解釋量。因此，在設計測驗時，可以藉由此三種試題成分的有無來操弄試題難度，而如此的結果更可應用於診斷測驗、題庫的建立、電腦化適性測驗以及動態評量的試題選擇上。

2.引用本研究結果應注意測驗內容與本研究所分析題材間的異同

由本研究中結果與討論一部分可知，本研究結果與過去的研究互有一致與不一致之處，而造成兩者不一致之處在於試題設計上的差異。故此，若欲引用本研究結果解釋數字系列完成作業試題難度的變異來源，應注意所分析之題材與本研究所使用的兩份測驗間的異同爲何，而不宜一概論之。

3.應用本測驗時應注意測驗使用的上限效應（ceiling effect）

研究者在針對數學測驗進行古典測驗理論試題分析時，發現本測驗試題難度對於高中二年級受試者有偏易的傾向，故對於高能力受試者能力的評量可能會受限制。因此，若欲應用數學測驗於學生能力的區別，可能對於中低能力學生較爲適當，而高能力學生可能較無法精確地評量其真正的能力。

(二)對未來研究的建議

1.以均衡的試題成分爲測驗設計取向

本研究所分析的數學測驗試題建構雖完全由電腦隨機命題，但在試題成分的分佈卻未盡理想。從表一有關數學測驗正式分析試題之試題成分分析統計可看出，八種試題特徵的題數分布並未達到均衡設計，是否有可能因此而影響整體試題難度的分布尚待進一步的探討。而後續研究者或可也採用以均衡的試題成分爲命題取向，以全面了解試題成

分對於試題難度的影響。

2.分析數字系列完成作業解題歷程中所可能涉及的認知歷程與解題策略

認知成分分析取向的缺點之一是各認知成分界定含糊不清，也因此研究者改採以可觀察與可操弄的試題成分爲本研究分析的對象。此外，認知成分分析取向的另一問題是該取向多以反應時間爲依變項，依此所建構的解題模式是否與以反應精確度爲依變項所建構之模式相同，尙待後續研究的證實。因此，後續研究者應可考慮以認知成分與解題策略爲自變項，反應精確度爲依變項，以了解數字系列完成作業可能涉及的認知歷程。

3.研究試題成分、受試者解題策略、解題認知成分、測驗間學習效應四者對於數字系列完成測驗試題難度的影響

本研究者認爲影響數字系列完成作業試題難度的因素除了試題成份外，尙且包含受試者解題策略、解題認知成分、測驗間學習效應等三種。以測驗編製者的立場而言，這四種因素中能操弄者爲試題成分，也因此本研究選取試題成分爲自變項。然而試題成分除了會影響試題難度外，尙且會影響解題認知成分、解題策略、以及測驗間的學習效應，甚且這些變項之間有可能也存在交互作用的關係。由於這些研究多偏向於認知心理學的實驗研究，本研究者因知識背景之不足而無法深入探討。然而爲求對數字系列完成作業的完整了解，後續研究者應可考慮加入這些變項，使測驗結果的解釋能有更實質的心理學意義。

4.發展二參數以及三參數成分分析的潛在特質模式

本研究係兼採用 LR 以及 LLTM 分析試題成分對於試題難度的影響，由於假定所取得之反應資料適配於 Rasch 模式，因此利用兩種分析方法皆能適用，所得之結果也相當一致。然而 Rasch 模式僅以難度參數描述試題特徵，對於反應資料諸多限制，反而無法了解不同反應資料的真實面貌，因此研究者認爲有必要發展二參數以及三參數成分分析的潛在特質模式，以供更多樣反應資料分析之用。

參考文獻

王寶墉（民 84）：**現代測驗理論**。台北：心理出版社。

郭生玉、林世華、蘇宜芬、宋德忠（民 87）：**高級中學資賦優異學生保送甄試升學鑑別用智力測驗及其題庫編製計劃第二年年度報告**。台北市：教育部中等教育司。

Butterfield, E.C., Nielsen, D., Tangen, K.L., & Richardson, M.B.(1985). Theoretically based measurements of inductive reasoning. In S.E. Embertson (Ed.)，*Test Design : Developments in Psychology and Psychometrics*, 77-147. New York：Academic Press.

Embretson, S. (1989). Program LINLOG: *An Extension of Fisher and Formann's Program for the Linear logistic Latent Trait Model.* Kansas: University of Kansas.

Green, K.E & Smith, R.M.(1987) A comparison of two methods of decomposing item difficulties. *Journal of Educational Statistics, 12(4)*, 369-381.

Holzman, T.G., Pellegrino, J.W. & Glaser, R. (1983). Cognitive variables in series completion. *Journal of Educational Psychology, 75(4)*, 603-618.

Kotovsky, K. & Simon, H.A. (1973) .Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology, 4*, 399-424.

LeFevre, J.A. & Bisanz, J. (1986). A cognitive analysis of number-series problems: Sources of individual differences in performance. *Memory & Cognition 14(4)*, 287-298.

Reckcase, M.D.(1979). Unifactor latent trait models applied to multifacto tests results and implications. *Journal of Educational Statistics, 4*, 207-230.

Simon, H. & Kotovsky, K. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review, 70(6)*, 534-546.

Sternberg, R. (1981). Intelligence and nonentrenchment. *Journal of Educational Psychology, 73(1)*, 1-16.

Sternberg, R.J. (1978). *Contrasting Conceptions of Intelligence and their Educational Implication. Technical Report No.14*. Office of Naval Research, Washington, D.C. ERIC Document Reproduction Service No: ED 166256.

Wilson, D., Wood, R.L., & Gibbons, R. (1983) .*TESTFACT: Test scoring and item factor analysis*[computer program] . Chicago : Scientific Software.

Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement, 24(4)*, 293-308.

收 稿 日 期：1999 年 3 月 11 日、
接受登刊日期：1999 年 5 月 26 日

Bulletin of Educational Psychology, 1999, 31(1), 139-165

The Study on Cognitive Component Analysis of the Number Series Completion Test

Shih-Hwa Lin

Department of Educational Psychology and Counseling National Taiwan Normal University Taipei, Taiwan, R.O.C.

Chia-Huei Yeh

Taipei Municipal Shih Pai Junior High School, Taiwan, R.O.C.

ABSTRACT

The purposes of this study were to decompose the difficulty level of number series completion task, and to examine if the constructs for detecting regular number series and irregular number series are different. The study was conducted with the pretest data of the math test in the Youth Intelligence Test, and focused on the test component analysis and the component parameter estimation that were calculated by using stepwise regression model and the Linear Logistic Latent Trait Model.

The major findings of this study are as follows:

1. The constructs to solve both regular number series items and irregular number series items are of the same dimension.
2. The major difficulty accountability of number series completion task comes from three components: if the series is moving or not, period of the series, and if the series is spurious identity or not. These three components significantly and positively related to the difficulty level of number series completion task.
3. In several cases, the interaction term among the components is able to account for the difficulty level of number series completion task.
4. The result shows that both stepwise regression and LLTM methods for estimation of component parameter are similar.

Keywords: number series completion test, cognitive component analysis