第三章　研究方法

本章主要先說明本研究之研究架構；其次，說明資料取得及描述，本研究採用參加國中基本學力測驗數學科的受試者之作答反應組型；另外，介紹所使用的研究工具，包括2001-2005年國中基本學力測驗數學科試題；在資料處理及分析部分，介紹如何將基本學力測驗數學科試題進行適切歸類、分析資料所需使用之模式及最適模式選擇之依據；最後說明本研究之研究程序。

第一節　研究架構

本研究旨在了解參加基本學力測驗數學科之國中生，其不同作答反應類型，在數學內容知識與數學認知能力方面的表現，其研究架構如圖3-1：

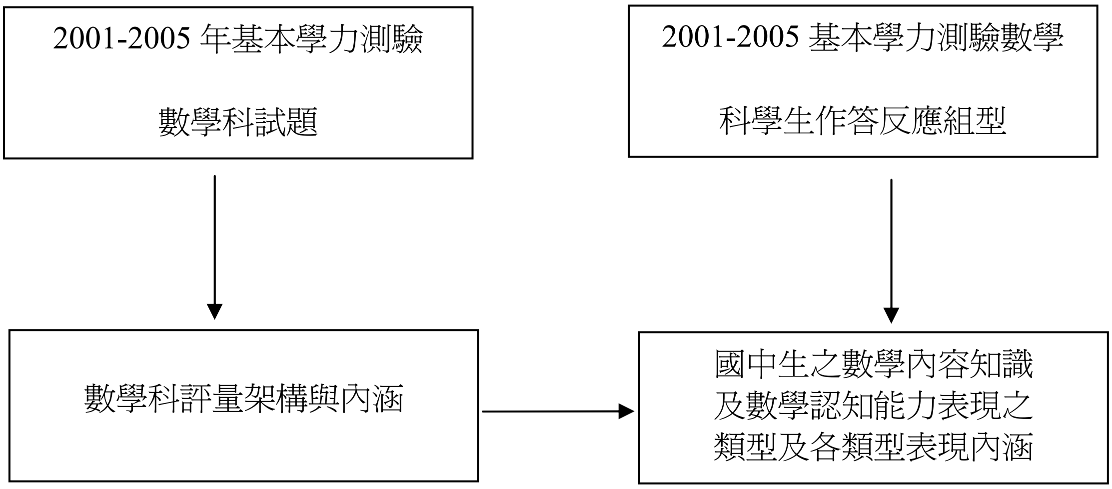


圖3-1：本研究架構圖

第二節　資料取得及描述

本研究之分析資料爲參加2001-2005年國中基本學力測驗之國中生，研究者經由向「國民中學學生基本學力測驗推動工作委員會」提出申請，獲得2001-2005年各年度第一次、第二次(共十次)國中基本學力測驗數學科的隨機抽樣樣本，每次各5000筆作答之反應組型，其中’0’表示答錯、’1’表示答對。

第三節　研究工具

本研究之研究工具爲2001-2005年第一、二次國中基本學力測驗數學科試題。基本學力測驗的命題方向，偏重在對學生未來的學習與生活都能有所幫助之基礎的、核心的、重要的知識與能力；命題概念是完整的、周延的，而非偏狹的、殘缺的。其特色如下(國民中學學生基本學力測驗推動工作委員會，2002)：1.標準化的；2.可比較的；3.一年多試、一試多用的；4.能力導向的；5.對教學有良性影響的。

國中生參加基本學力測驗數學科測驗反應之描述性統計，如表3-1，其中各次測驗之內部一致性係數，係爲研究者所申請之5000筆受試者反應資料所得：

表3-1：基本學力測驗數學科測驗反應之描述性統計表

年度(次別)	總人數				總題數	最低分	最高分	平均答對率	內部一致性係數
	有效人數	點字卷人數	缺考人數	違規不計分暨撤銷資格人數					
2001(1)	302388				32	0	32	0.579	0.893
	299368	0	3010	10					
2001(2)	176414				31	0	31	0.597	0.905
	167440	0	8970	6					
2002(1)	299714				31	0	31	0.514	0.893
	295917	19	3775	3					
2002(2)	189622				31	0	31	0.610	0.913
	184663	2	4953	4					
2003(1)	313239				31	0	31	0.631	0.924
	307194	14	6028	5					
2003(2)	183436				31	0	31	0.688	0.919
	177078	0	6355	3					
2004(1)	314675				32	0	32	0.612	0.934
	311481	18	3171	5					
2004(2)	188783				32	0	32	0.688	0.905
	179630	0	9151	2					
2005(1)	322330				33	0	33	0.642	0.927
	319245	12	3069	4					
2005(2)	192372				33	0	33	0.640	0.876
	182406	0	9965	1					

綜合2001-2005年基本學力測驗數學科試題反應之描述性統計，可得以下四點結果：

一、就參考考試的人數：參加每年度第一次基本學力測驗的總考生約爲30萬人左右，而參加每年度第二次基本學力測驗的總考生則約爲18萬人左右。

二、就數學科各次試題題數：試題之總題數介於31到33題之間。

三、就試題平均答對率：2001年第一次、2001年第二次、2002年第一次，這三次的試題平均答對率約爲0.5~0.6之間；自2002年第二次至2005年第二次共七次的考試，試題的平均答對率約爲0.6~0.7之間。

四、就各次之隨機樣本5000筆資料，進行內部一致性分析，得到內部一致性係數約在0.9左右，表示測驗試題之一致性很高。

第四節　資料處理及分析

本節首先說明將各次基本學力測驗數學科試題進行適切歸類之程序；其次說明試題分析之資料處理及分析，包括模式單向度考驗、混合Rasch模式模式分析及國中生在數學內容知識與數學認知能力之試題答對率分析。

一、各次基本學力測驗數學科試題之歸類程序

(一)參與研究之國中現職數學教師

爲使試題歸類具一致性，研究者乃邀請三位現任國中數學教師參與研究，以下則說明三位教師之背景及試題歸類之程序及結果，如表3-2：

表3-2：參與研究之國中現職數學教師背景資料表

背景資料 / 教師	學歷	任教國中數學年資
A教師	大學	3年
B教師	大學	6年
C教師	碩士	4年

(二)試題歸類之程序

為求試題歸類之一致性，本研究之試題歸類程序如下：

1.在正式進行試題歸類前，首先邀請參與研究之數學教師，針對本研究有關之數學內容知識與數學認知能力進行文獻回顧，並請各數學教師界定其分類，其次，由研究者彙總各教師訂定之數學內容知識與數學認知能力之分類，並經由與各數學教師之共同討論，訂定本研究之分類：即數學內容知識分為「數與量、幾何與空間概念、資料分析統計與機率、代數」、數學認知能力分為「概念理解、程序知識與執行、問題解決」。

2.針對本研究所界定數學內容知識與數學認知能力，與參與研究的三位數學教師進行溝通，達到歸類定義之一致性。

3.在數學內容知識之分類確立後，與參與研究之數學教師，針對基本學力測驗數學科各試題所屬之學習單元進行討論，形成數學學習單元分類之依據。

4.就2001-2005年共十份試題，由研究者與參與研究的三位教師，共四人，分別就各試題所屬之數學內容知識、數學認知能力及數學學習單元，個別進行試題歸類，每份試題至少有兩人進行歸類。

5.在個別進行歸類之後，由研究者進行歸類之彙總，同時，找出歸類上出現歧異的

試題。

6.研究者與三位參與研究之教師，針對歸類上出現歧異之試題進行討論，表示想法與意見，在出現不同意見時，則重新檢視先前對數學內容知識與數學認知能力的定義。

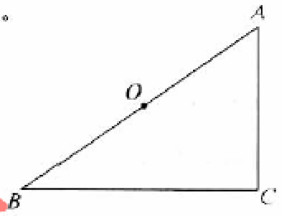
7.經由充分討論，形成歸類之共識。

以下則列舉例子，說明如何形成歸類共識：

例一：

2001年第一次測驗第24題

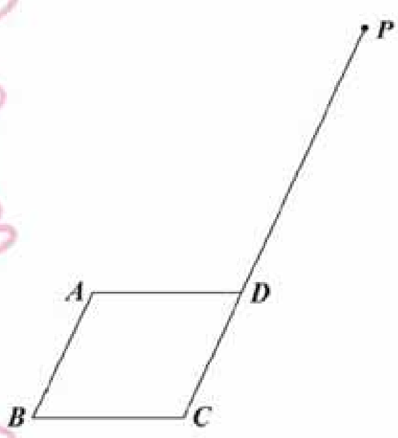
24. 如圖(十一)，已知在△ABC中，$\angle ACB=90^\circ$且$\overline{BC}>\overline{AC}$。
求作：一圓與$\overline{AC}$、$\overline{BC}$相切，且圓心O在AB上。
下列四個取得圓心O的作圖方法，何者正確？
(A)取 AB 中點為 O
(B)作 AC 中垂線交 AB 於 O
(C)作 BC 中垂線交 AB 於 O
(D)作$\angle ACB$平分線交 AB 於 O



圖(十一)

2005年第一次測驗第30題

30. 如圖(二十)，四邊形$ABCD$為一平行四邊形，P在直線CD上，且$\overline{PD}=2\overline{DC}$。甲、乙兩人想過$P$點作一直線，將平行四邊形分成兩個等面積的區域，其作法如下：
甲：取$\overline{AD}$中點E，作直線PE，即為所求。
乙：連接$\overline{BD}$、$\overline{AC}$交於O，作直線PO，即為所求。
對於甲、乙兩人的作法，下列判斷何者正確？
(A)甲、乙皆正確
(B)甲、乙皆錯誤
(C)甲正確，乙錯誤
(D)甲錯誤，乙正確



圖(二十)

上述兩題皆屬於作圖題，在試題歸類中，可初步歸類至「作圖證明」，然而2001年度第一次測驗第24題的作圖牽涉到「中線、角平分線、中垂線之性質」，由於2001-2005年的測驗中，此類試題有若干題，因此在參與研究之教師經由討論後，決定將牽涉到「中線、角平分線、中垂線之性質」之試題歸爲一類，而其它的作圖題則歸類至「作圖證明」。

例二：

2001年第一次測驗第14題

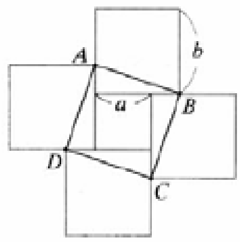
14. 將一塊邊長為 a 的正方形，與四塊邊長為 b 的正方形（其中 $b>a$），拼成如圖(四)，其中 $\overline{AB}$、$\overline{BC}$、$\overline{CD}$、$\overline{AD}$ 形成一個四邊形，則四邊形 $ABCD$ 的面積為多少？

(A) $b^2+(b-a)^2$

(B) b^2+a^2

(C) $(b+a)^2$

(D) a^2+2ab



圖(四)

本題在學習單元屬於「商高定理」，在數學認知能力方面，一教師認爲應屬於「程序知識與執行」，另一教師則認爲應屬於「概念理解」，經討論後，認爲本題是測驗學生是否了解商高定理的基本意義，並未牽涉到計算，因此最後一致同意將試題所屬之數學認知能力歸類至「概念理解」。

例三：

2004年第一次測驗第1題

1. 已知甲 $=4\frac{3}{8}$、乙 $=4\times\frac{3}{8}$、丙 $=4+\frac{3}{8}$，比較甲、乙、丙三數的大小，下列敘述何者正確？

(A) 甲＝乙

(B) 甲＝丙

(C) 甲＜乙

(D) 甲＜丙

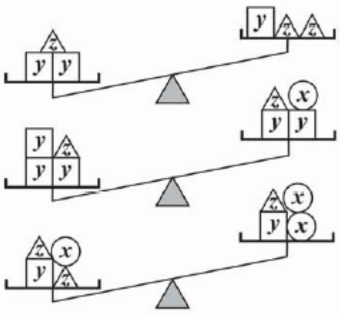
本題在學習單元屬於「(分數)四則運算」，在數學認知能力方面，一教師認爲

應屬於「程序知識與執行」，另一教師則認爲應屬於「概念理解」，經討論後，認爲本題是測驗學生是否了解四則運算的基本意義，並未牽涉到計算，因此最後一致同意將試題所屬之數學認知能力歸類至「概念理解」。

例四：

2005年第一次測驗第9題

9. 圖(二)是將積木放在等臂天平上的三種情形。若一個球形、方形、錐形的積木重量分別以 x、y、z 表示，則 x、y、z 的大小關係為何？
(A) $x>y>z$
(B) $y>z>x$
(C) $y>x>z$
(D) $z>y>x$



圖(二)

本試題之數學內容知識屬於「代數」，在細分學習單元時，一教師認爲是「等量公理」，另一教師則認爲應屬於「不等式」，經討論後，認爲此試題屬於「不等式下的等量公理」，因此歸類於「不等式」。

二、試題分析

(一)模式單向度考驗

由於混合Rasch模式結合潛在類別分析與項目反應理論，因此在進行分析前，需先針對項目反應理論之基本假定「單向度」進行考驗，本研究利用因素分析方法，針對2001-2005年基本學力測驗數學科受試者作答反應組型(共十次測驗)進行分析，參考Reckase(1979)所提出的單一主要向度標準(λ_1佔總變異的20%以上)，以及第一特徵值(λ_1)與第二特徵值(λ_2)之比大於4，作爲判定單一向度的標準。

(二)混合Rasch模式分析

本部份主要是在說明如何選擇適合之模式以描述資料，其方式有兩種，分別

爲：比較觀察值與期望值、比較不同模式之間的相對適合度。陳述如下：

1.比較觀察值與期望值

直接比較觀察值和預測值的取向上，常使用的兩種統計量爲Pearson χ^2 及Likelihood G^2(Agresti, 1996)，Cressie和Read也提出一個在χ^2和G^2之間的CR統計量(Cressie & Read, 1984; Read & Cressie, 1988) 對於這三種考驗統計量，較小數值的統計量與其自由度有關，對應較大的p值，表示有資料與模式之適合度較好。

此取向的一個特例爲，當樣本大小與估計變數的數目相比是相對較小的，即期望次數在變數的一些可能的組合下小於5或是0，我們稱此種資料爲稀少的資料(sparse data)，因此前述所提的考驗統計量不會有很好的效果，並且所獲得的模式適合度的p值是不能相信的(Dayton, 1988)。

針對非常稀少(sparse)的資料，如Pearson χ^2 或Likelihood G^2之適合度統計量，並不能經由卡方分配來估計(Collins, Filder, Wugalter, & Long, 1993; Koehlen & Larntz, 1980)，此時需使用參數的Bootstrap法(Efron & Tibshirani, 1993)，以藉由漸近的(asymptotic)分配來考驗適合度的統計量。von Davier(1997)指出當資料是非常稀少的，Bootstrap法在Pearson χ^2 或CR統計量爲好，但在Likelihood G^2和Freeman-Tukey統計量，Bootstrap方法則並未較好。當資料不是稀少時，不論是Bootstrap法或卡方考驗，上述四種統計量，Pearson χ^2 、Likelihood G^2、CR統計量、Freeman-Tukey統計量，皆可用來作爲模式模式資料---適合度的考驗。

Bootstrap法應用電腦對研究者蒐集來的資料進行重複抽樣，再以實徵結果估計某特定統計量的標準誤或抽樣分配(盧雪梅，民88)。Bootstrap法又分爲無參數的bootstrap法(nonparametric bootstrap)及參數的bootstrap法(parametric bootstrap)(Efron & Tibshirani, 1993)。無參數的bootstrap法，其方法爲針對未知分配的母群體進行隨機置回的抽樣(random sampling with replacement)，直到抽足N個樣本，稱爲一個bootstrap樣本，反覆進行此程序，直到獲得研究者預定的bootstrap樣本數，進行統

計分析，以估計研究者感興趣的統計量之標準誤或抽樣分配；而參數的bootstrap法，是從一個已知分配的母群體中，以隨機置回的抽樣，抽取N個樣本，反覆進行此程序，直到獲得研究者預定的bootstrap樣本數，利用各樣本估計母體參數，以獲得有興趣之統計量。實務上，參數的bootstrap法通常先從未知分配的母群中隨機抽取一組樣本，估計此樣本中研究者有興趣的統計量，同時視此統計量爲樣本之參數；其次，針對已假定參數的前述樣本，以隨機置回的方式，進行重複抽樣，一直到研究者所需的樣本數爲止；最後針對各bootstrap樣本，進行統計量之估計，因而可對參數進行相關的統計推論。

以 WINMIRA 軟體進行混合 Rasch 模式之分析，若分析的資料爲非常稀少的資料，則可進行 bootstrap 法進行模式適合度之考驗，在 WINMIRA 軟體之分析，所使用的 bootstrap 法爲參數的 bootstrap 法。

2.比較不同模式之間的相對適合度

此取向是比較不同模式之間的相對適合度，在此取向中有兩種情況，第一種情況是當兩個模式的比較是巢狀的(nested)，亦即一個模式是在另一個模式之中，這兩個模式是巢狀的，並且可以在限制一個或多個參數的情況下獲得一個介於這兩個模式之間的較簡單的模式，然而在潛在類別分析的模式中，不同的類別之間並非是巢狀的(Dayton, 1988; McCutcheon, 1987; Vermunt & Magidson, 2002)；亦即第二種的情況是針對模式之間的適合度的比較，在此種情況下使用G^2通常不會有好的效果，我們通常使用訊息的規準如Akaike Information Criteria(AIC)、Bayesian Information Criteria(BIC)或CAIC(Vermunt & Magidson, 2002)，這三種訊息規準背後的概念是一樣的，分述如下(Uebersax, 2003)：

(1)AIC值

此訊息規準是針對參數估計的數目進行L^2的校正，其計算公式如下：

$$AIC = -2\ln(L)+2P$$

ln(L): 對數概率

P: 估計模式參數的數目

(2)BIC值

此訊息規準是針對參數估計的數目和樣本大小進行L^2的校正，在三個訊息規準值中，是最好的規準值，其計算公式如下：

BIC = -2ln(L)+P×ln(N)

ln(L): 對數概率

P: 估計模式參數的數目

N: 觀察的總個數

(3)CAIC值

其計算公式如下：

CAIC = -2ln(L)+P×ln(1+ln(N))

ln(L): 對數概率

P: 估計模式參數的數目

N: 觀察的總個數

對於模式具相同層次的適合度，具較少參數的模式是較好的訊息規準，較小的值通常表示模式的適合較好，通常BIC是較好的訊息規準，學者指出BIC是較一致性的(Li & Nyholt, 2001)，以及傾向於較AIC爲較節儉的(parsimonious)模式(Lin & Dayton, 1997)。

3.模式適合度與模式選擇的參考依據---結合卡方值及訊息規準值

假設第一類型錯誤爲.05，參數的bootstrap之p值小於.05，表示潛在類別之資料與模式的符合度是不適切的；而參數的bootstrap之p值大小於.05，表示潛在類別之

資料與模式是適切的。

然而，若有若干個模式之參數的bootstrap之p值大於.05時，此時並非是p值最大者的模式即爲最適合之模式，因爲參數的bootstrap之p值不能作爲模式適合度之測量，原因爲此時之p值只顯示資料與模式之間的差異是非顯著的，因此需使用其他判斷規準來更進一步的選擇最適切的模式。

因此，在有若干個模式之參數的bootstrap的p值大於.05時，我們可以說這些模式之資料與模式之間的差異是非顯著的，此時，我們必須同時考慮訊息規準值，一般來說，訊息規準值中最小的BIC值所屬之模式，即爲最適合的模式。

4.以WINMIRA軟體進行混合Rasch模式分析

在以WINMIRA軟體(von Davier, 2001)進行受試者作答反應組型之分析前，需先根據研究的需要，設定以下項目：

1.設定模式：WINMIRA共提供三種分析模式，分別爲次序變項之潛在類別分析(The latent class analysis for ordinal variables)(Rost, 1988)、混合Rasch模式、混合模式(Hybrid Model)(Yamamoto, 1987)，本研究所使用的分析模式爲混合Rasch模式

2.設定潛在類別數目：首先設定由一個潛在類別的分析開始，在WINMIRA中，一個潛在類別的分析即爲傳統的Rasch模式(RM)，接著進行二個、三個……等潛在類別的分析，直到符合解釋資料的潛在類別數爲止。

3.若分析的資料屬於稀少資料，亦即樣本大小與估計變數的數目相比是相對較小的，同時期望次數在變數的一些可能的組合下小於5或是0時，此時則須選擇bootstrap法進行卡方值考驗，了解資料與模式之適切度。

在上述分析模式設定、並針對資料進行分析後，可得到研究所需分析結果：

(一)期望類別次數

本研究所獲得之基本學力測驗數學科學生作答反應組型，爲二元計分，其中’1’

表示答對、’0’表示答錯。期望類別次數(expected category frequencies)即爲每一試題理論上0或1發生的次數，若將答對次數除以總次數，即可得到試題理論上的答對率。

(二)判斷模式適合度之卡方值

若資料爲稀少資料，則須選擇bootstrap法進行卡方值(chi-square)考驗，了解資料與模式之適切度。

(三)訊息規準值

不同模式之間相對適合度的另一種方法爲訊息規準值(information criteria)，WINMIRA提供AIC、BIC、CAIC三種訊息規準值，其中訊息規準值中最小的BIC值所屬之模式，即爲最適合的模式。

(四)各潛在類別佔母群體之比例

進行潛在類別分析時，根據受試者作答反應組型進行歸類，計算各潛在類別所佔全體受試者的比例，比例計算的方式爲：歸類到某一潛在類別的受試者數目，除以全體受試者數目。

三、國中生在數學內容知識與數學認知能力之試題答對率分析

以SPSS for Windows 13.0統計軟體，針對各作答反應類型學生之數學內容知識及數學認知能力進行分析：

(一)針對2001-2005年各次測驗之二元計分之Rasch模式分析所得之試題答對率，分別計算數學內容知識(數與量、幾何與空間概念、資料分析統計與機率、代數)與數學認知能力(概念理解、程序知識與執行、問題解決)之試題平均答對率。

(二)針對2001-2005年各次測驗之混合Rasch模式分析所得之試題答對率，分別就最適模式爲三組之六次測驗及最適模式爲四組之四次測驗，分別計算各組在數學內容知識與數學認知能力之試題平均答對率。

第五節　研究程序

本研究之研究程序如下：

一、蒐集相關資料及文獻撰寫研究計畫

自民國九十四年二月起，開始著手閱覽蒐集相關文獻，同年六月，確定研究主題，開始撰寫研究計畫，計畫提報指導教授審核修訂後，正式定名爲「國中生數學內容知識與數學認知能力之混合Rasch模式分析研究」。

二、申請國中畢業生參與基本學力測驗數學科之資料

向國民中學學生基本學力測驗小組申請2001-2005年，各次國中基本學力測驗數學科的抽樣樣本各5000筆作答反應組型。

三、基本學力測驗數學科試題之歸類

邀請國中現職數學教師參與研究，共同擬定本研究之數學科評量架構，並進行基本學力測驗數學科試題之歸類。

四、資料處理及分析

以WINMIRA軟體，針對所申請之基本學力測驗數學科學生作答反應組型進行混合Rasch模式分析；統整數學科試題之歸類及混合Rasch模式分析之結果，以SPSS for Windows 13.0統計軟體進行研究所需之資料分析。

五、歸納研究結果，撰寫論文

最後將各種分析結果加以歸納、解釋，並整理彙總各部份資料，撰寫並提出完整之研究論文。