

第三章 研究方法

本章共分四節，依次說明本研究之研究架構、研究樣本與工具、實施流程與研究設計、資料分析與統計處理等四大部份。

第一節 研究架構

本研究首在以最大測驗訊息量法為精熟標準設定的核心，形成如圖 3-1 所示之研究架構。首先，由文獻探討中歸納出「以能力標準設定為核心的測驗建構流程」中所需的基本元素，以作為本研究架構的成分，同時，由過去研究中方法與議題的演進，歸納出「方法中元素組合」與「過程檢定」的訊息，兩者相加結合，強調著以廣義測驗發展的過程觀點檢視本研究所運用最大測驗訊息量法。探討其理論性，並尋求最佳測驗結果的解釋方法，以提供多元的效度證據，此外，探究主要影響分類一致性信度的測驗長度與異質性議題，評估其與上述方法間的互動、影響效果，期能發展出一套適切融合各測驗發展層面的精熟標準設定方法，最後，提出論文報告。

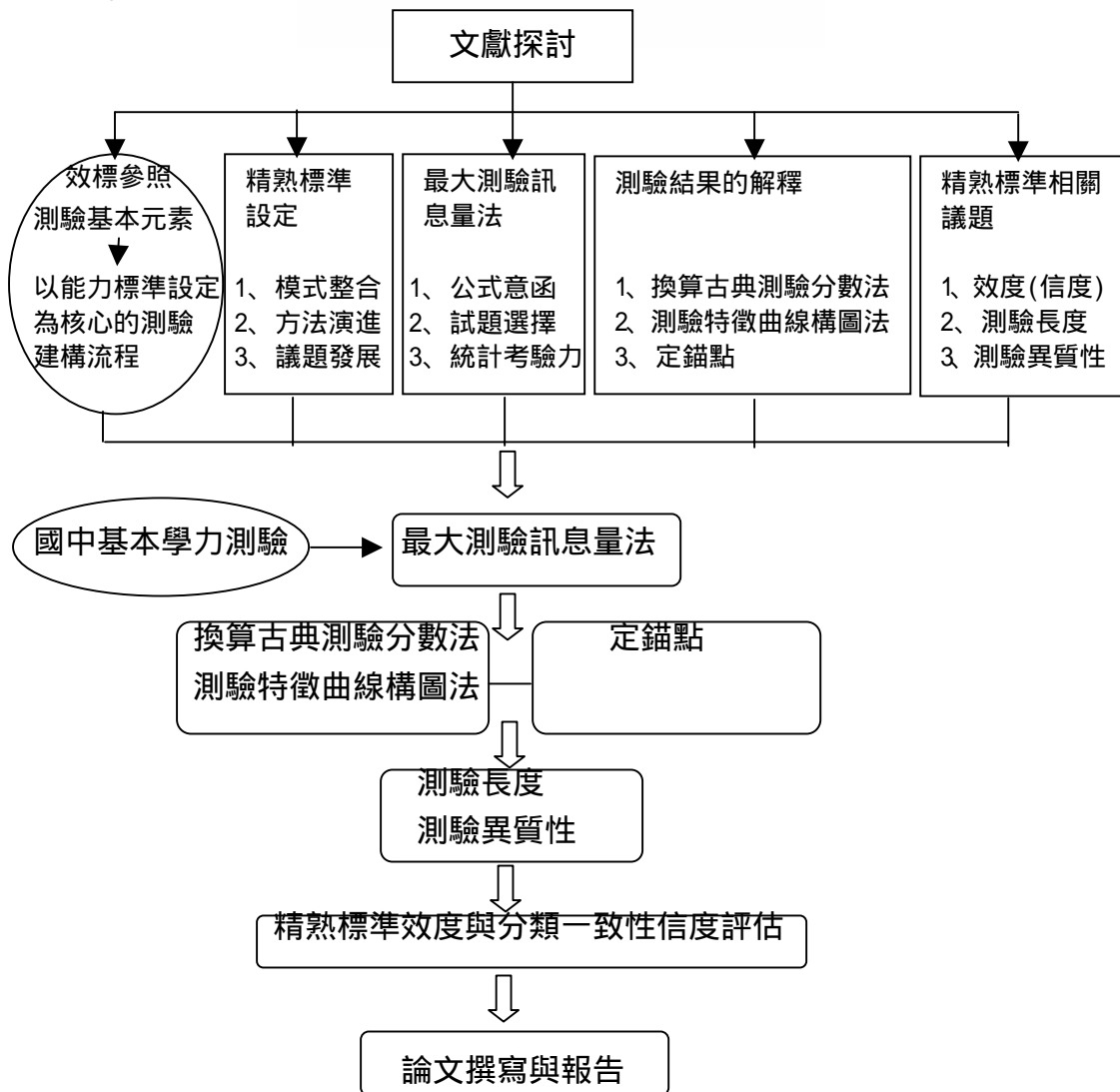


圖 3-1 本研究研究架構圖

第二節 研究樣本與工具

壹、研究樣本

本研究對象乃以報名參加 2002 年國中基本學力測驗者中，從全國 18 個考區中隨機抽取兩次測驗皆有參與的考生共 20000 人(本項資料數據是由國中基本學力測驗推動小組，從全國約 30 萬名考生資料庫中，隨機抽取 2 萬名而得)，其相關人數分配如表 3-1、3-2 所示。而在考量相關因素後(於「研究工具」內討論和說明)，最後決定只針對自然科的兩次測驗資料進行分析，以作為本研究之分析樣本。

表 3-1 國中基本學力測驗第一、二次測驗男女生受試者分配表

性別	人數	百分比(%)
男	10519	52.595
女	9481	47.405
總計	20000	100.000

表 3-2 國中基本學力測驗第一、二次測驗各區域受試者分配表

區域	人數	百分比(%)
北	8308	41.540
中	6075	30.375
南	5081	25.405
東	387	1.935
離島	149	0.745
總計	20000	100.000

貳、研究工具

本研究乃以探討整個精熟標準設定流程為主軸，而測驗工具的選擇則為重要開端，其不僅需符合效標參照測驗基本定義，又需配合本研究所運用的方法、研討議題，且能融入整個精熟標準設定模式中，並藉此以提供多元效度的過程證據，而在考量各層面因素後，最後選擇國中基本學力測驗作為本研究研究工具，茲簡述如下：

一、國民中學學生基本學力測驗

國中基本學力測驗的起源是來自 1996 年行政院教育改革委員會總諮議報告書(行政院教育改革委員會, 1996)，此報告書中提到，基本學力測驗的命題方向，將偏重對學生未來的學習與生活有所幫助之基礎的、核心的、重要的知識與能

力。另教育部頒訂 90 學年國中基本學力測驗的宗旨，在評量國民中學學生基本能力表現及其發展潛能，由此視之，雖然國中基本學力測驗在解釋分數時是以常模參照測驗的角度(國民中學學生基本學力測驗推動工作委員會，2003)，但其編製試題的出發點，因為是用以測量學生基本能力，作為能力檢定的考試，仍是採用「效標參照測驗」的觀點。因此，本研究在考量該工具本質上的合理性、適切性與接續探討之研究期望，希望藉此將結論推引至其他資格檢定考試上，所以，選擇國中基本學力測驗作為本研究之工具，並希望讀者能將它視為是一種具備優良品質的「基本能力檢定測驗」工具。

就文獻探討中「效標參照測驗」的定義綜整結果，可從中萃取出幾項元素：明確定義的內容範圍、學習目標(核心目標與具體能力指標)、精熟標準、能力解釋。同時，亦代表著整個精熟標準設定流程，在此，檢視條件組成的完備性，首先，就內容範圍、學習目標簡述之。

(一)內容範圍

國中基本學力測驗取材範圍源自八十三年十月教育部公告修正發佈之「國民中學課程標準」為規範，包含有國文、英語、數學、認識台灣、歷史、地理等二十餘種科目，最後，整體歸納而為國文、英語、數學、社會學科與自然學科等五大領域(教育部，1998)。實際取材施測上，則由國民中學學生基本學力測驗推動工作委員會詳細界定測量的內容範圍，就 91 學年國中基本學力測驗兩次測驗為例，命題範圍為五大學科，從一年級上學期至三年級下學期第二次段考前之進度，就自然科而言，其內容指標包含生物體的組成、生命現象、物質性質的探討、化學反應、宇宙的組成與變動、能量的形式與轉換、環境與健康、力與運動、生態平衡等九項。詳細細則可參考國民中學學生基本學力測驗推動工作委員會(國民中學學生基本學力測驗推動工作委員會，2002a、2002b)。

(二)學習目標

學習目標象徵著整份測驗的核心，其不僅架構測驗試題的規劃，亦隱含著決策者對於最低基本能力者的界定，具體而言，即是所企求的精熟標準點位置。以下茲就國中基本學力測驗的核心目標與具體能力指標作簡摘。

1、核心目標

教育部中等教育司於 1998 年 6 月成立「國中基本學力指標建構小組」著手建構國中基本學力指標。研究小組採用目標分析、教材分析、意見座談與小組討論等方法建構基本學力指標。其中，研究小組界定所謂的「基本」，就層次而言乃是指基礎、核心、重要的，而非高深、外圍或細微末節的；就範圍而言，是指完整周延的，而非偏狹或殘缺的。就自然科而言，包含健康教育、生物、地球科學、理化等科目，各科皆有其既定核心理念，如理化科則強調增進學生對物質和能量世界的瞭解，以建立理化基本知識與概念，養成具有科學素養的國民(教

育部，1998，p67)。

2、具體能力指標

在各科核心目標下，依布隆姆(Bloom, B. S.)有關教學目標階層分類方式，進一步依據內容指標，製訂具體能力指標，兩者相互搭配下，以作為試題的編製，更為往後試題取樣奠基。以自然生物科為例，其階層可為：科學基本能力 → 探索的過程 → 察覺特徵 → 能從所給予的生物活體或圖片指出表現的生命現象(教育部，1998，p69)。

就整個精熟標準設定流程中，國中基本學力測驗已具備良好開端，以下進一步就其本研究所採科目的測驗品質、考量因素詳述之。

二、2002 年國中基本學力測驗自然科兩次測驗試卷

國中基本學力測驗乃教育部自民國 90 年起開始採行，以替代傳統聯考的測驗方式，建立於 IRT 理論模式下，採二元計分，用以檢測學生基本的學科知識，其編製係依照標準化測驗的編製程序(涂柏原、陳柏熹、章舜雯、林世華，2000)而來，從各學科領域雙向細目表的擬定、試題的編製與取樣、預試、試題分析以及信、效度的建立，歷經多位學者專家修訂及數年努力而成，具有極高的測驗品質，可用以作為本研究之工具。而本研究之所以會選擇自然科測驗科目，主要是考慮下列因素：

(一)研究控制取向

為了使測驗中受試者能力分配與測驗異質因素干擾減至最低，乃先針對 5 個考科的兩次測驗，經相依樣本 t 考驗後，儘量選取受試者在各科兩次測驗上平均得分差異最小(以此作初步難度比對)與偏態值相近的科目為分析資料。

(二)研究操弄取向

科目需具備一定測驗長度，不然無法進行測驗長度因素的操弄。

(三)測驗品質

考量該科目測驗是否具一定信、效度與符合 IRT 理論架構。

在考量上述因素後，自然科兩次測驗資料最接近條件需求，其結果如表 3-3 所示，受試者在兩次測驗上平均分數差異雖然達顯著，此乃因為樣本人數過大所造成的影響結果，其實質上的差異只有 0.73，仍在可接受範圍內。而在測驗長度、 α 信度值與偏態差異上也符合一定要求水平，此外，測驗品質方面，參照謝進昌、余民寧(2005)以 Mantel-Haenszel 法檢測 91 年度國中基本學力測驗不同考區可能產生的差異試題功能(differential item functioning, 簡稱 DIF)，發現自然科兩次測驗雖有些許 DIF 現象，但多屬輕微程度，幾乎可予忽略；而在 IRT 單向度假設上，經因素分析結果如表 3-4、表 3-5 所示，第一次測驗與第二次測驗的最大特徵值乃分別為第二大特徵值的 7.31 與 7.28 倍，而第二大特徵值

分別為第三大特徵值的 1.42 與 1.64 倍，大致符合 Lord(1980)所提單向度考驗的標準。綜合上述考量，本研究於是決定選取自然科兩次測驗資料作為實徵分析對象。

表 3-3 自然科兩次測驗各指標摘要表

自然科	題數	平均分數(標準差)	差異	標準化 α 值	偏態係數	單向度假設
第一次	58	35.09(11.93)	0.73**	0.9272	-0.003	符合
第二次	58	34.36(13.20)		0.9432	-0.233	符合

註：**表 $p < .01$

表 3-4 自然科第一次測驗試題因素分析摘要

主成分	特徵值	解釋總變異量%	累積解釋總變異量%
1	12.053	20.780	20.780
2	1.650	2.845	23.625
3	1.165	2.009	25.634
4	1.124	1.937	27.571
5	1.047	1.805	29.376
6	1.011	1.743	31.119

表 3-5 自然科第二次測驗試題因素分析摘要

主成分	特徵值	解釋總變異量%	累積解釋總變異量%
1	14.247	24.564	24.564
2	1.957	3.374	27.938
3	1.196	2.061	30.000
4	1.096	1.890	31.890

三、2002 年國中基本學力測驗自然科兩次測驗作答結果資料

本研究擬分析 20000 筆考生作答資料，資料全部來自教育部資助之「國民中學學生基本學力測驗研究發展計畫」(該計畫係由台灣師範大學心理與教育測驗研究發展中心執行)，並由該計畫的「國民中學學生基本學力測驗推動工作委員會」所釋出。作者感謝上述機構及人員提供資料協助，然本文內容分析所得結果與結論，全由作者自行負責，與該機構及人員無關。

四、SPSS 統計軟體與 BILOG-MG

本研究擬採用之工具，分別為 SPSS(Statistical Package for Social Science)統計軟體與 IRT 軟體 BILOG-MG 程式 3.0 版(Zimowski, Muraki, Mislevy, & Bock, 2003)，這兩套軟體自早期發展至今，歷經數年研發、修訂，其分析和估計效果是值得肯定的。

第三節 實施流程與研究設計

本研究實施流程主要關注於初步資料整理與科目篩選方式，同時描述最大測驗訊息量法、換算古典測驗分數法、轉換分數效益評估、測驗特徵曲線構圖法與定錨點的執行流程並探討測驗長度、異質性等相關議題之研究設計，茲說明如下：

壹、資料整理與科目篩選

一、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，取得 20000 筆分析資料，經整理後，資料格式如表 3-6 範例所示。

表 3-6 自然科第一次測驗第 1 位受試者作答資料範例

1	1111010100000000000100100111100001000000011111100000000110 1 5
註：1-5 流水號；7-64 受試者於各題作答結果；66 性別代碼；68-69 考區代碼。	

二、針對原始資料，經初步整理，針對上述科目篩選因素，進行分析，採用 SPSS 統計軟體計算各科測驗的描述統計量數、內部一致性信度、相依樣本 t 考驗、因素分析等，篩選結果乃決定以自然科兩次測驗為本次研究對象。

三、有藉於本研究探討面向較為複雜，因而，為統一與簡化研究結果，在此，乃針對資料作些許調整。首先，就獲得之 20000 名考生資料，利用 SPSS 軟體，依據各考區人數比例，將其隨機分派成兩組，以其中一組考生的作答反應，將之稱為「試題產生樣本」，採用 BILOG-MG 軟體，求得兩次測驗試題參數，形成小型題庫，以提供不同研究議題中試題參數來源之用（相關試題參數請見附錄一）。之後，搭配另一組受試者資料，將之稱為「研究操弄樣本」，以作為各研究議題操弄時的基礎。該各 10000 名考生相關描述統計資料如表 3-7 所示，於兩次測驗上平均分數差異雖達顯著，但實質差異僅為 0.76，實屬微小，而偏態係數差異亦在合理範圍，受試者於兩次測驗上之表現可謂穩定，且 α 信度值、單向度假設上也符合一定要求水平，以利 IRT 的分析。

表 3-7 自然科兩次測驗研究操弄樣本各指標摘要表

自然科	平均分數(標準差)	差異	標準化 α 值	偏態係數	單向度假設
第一次	35.15(11.93)	0.76**	0.9273	-0.023	符合
第二次	34.39(13.23)		0.9435	-0.239	符合

註：**表 $p < .01$ ；單向度結果與本章第二節研究工具檢測驗結果一致，故不再加以引述。

貳、最大測驗訊息量法

對於最大測驗訊息量法的實施，首先，利用試題產生樣本求得之已知試題參數，分別計算研究操弄樣本 10000 名考生之能力估計值與兩次測驗的最大測驗

訊息量對應之 θ_{\max} ，即代表最大測驗訊息量法於該次測驗的精熟標準。在進行第一次分類時，乃以第一次測驗求得之 θ_{\max} 為精熟標準，將受試者能力值大於/等於此 θ_{\max} 者，視為精熟者；反之，小於此 θ_{\max} 者則視為未精熟者，而以第二次測驗作為效標，以求出精熟者與未精熟者在效標上的分類一致性。另在進行第二次分類時，則以第一次測驗為效標，改採用第二次測驗求得之 θ_{\max} 為精熟標準(依照同樣步驟區別出精熟/未精熟者)，再計算其在效標上的分類一致性。在如此交叉驗證下，進行百分比一致性與 κ 係數的計算，以確立根據最大測驗訊息量法分類成精熟/未精熟者的一致性程度。

參、換算古典測驗分數法

對於換算古典測驗分數法的實行，概念上乃針對自然科兩次測驗研究操弄樣本，分別計算其古典測驗的答對題數總分，再利用余民寧、汪慧瑜(2005)所提轉換 IRT- 能力值的概念，針對受試者實際表現進行分數轉換。執行步驟如下所述：首先，篩選出研究操弄樣本 10000 名受試者中總分 1 分的人(理論上約有 $\binom{58}{1}$ 種組合的人數，參見公式 2-12 所示)，計算其平均 IRT- 能力值(即 $\bar{\theta}_1$)，即視為古典測驗分數 1 分時所對應的 IRT- 能力值，再依次分別計算總分 2、3、4、58 分與其平均 IRT- 能力值的對照值(即 $\bar{\theta}_2$ 、 $\bar{\theta}_3$ 、 \dots 、 $\bar{\theta}_{58}$)。

如此，將該次測驗最大測驗訊息量所對應之 θ_{\max} 求出後，即可對照求得 θ_{\max} 值相當於古典測驗答對題數的總分大約是幾分(或幾題)，並以此分數作為 IRT- 能力值所對應的古典測驗通過分數，再將每位考生的古典測驗原始答對題數總分大於或等於此通過分數者，即視為精熟者，低於此通過分數者，則視為未精熟者。接續，再進行如上述之兩次交叉驗證分類，同時進行後續分類一致性的指標計算。

肆、測驗特徵曲線構圖法

對於測驗特徵曲線構圖法的轉換方式，實行時乃採以公式直接進行分數轉換。首先，將最大測驗訊息量法求得之 θ_{\max} 視為某考生之能力值，代入三個參數對數形模式之正確反應機率數學函數(試題參數由上述已知參數代入)，即可求得此於每一試題之答對機率值，如此，再將各試題答對機率值加總，可得轉換後之測驗特徵曲線答對機率值，將此視為測驗特徵曲線構圖法求得之轉換後古典測驗答對題數，即作為作為最大測驗訊息量法求得之 IRT 精熟標準 能力值所對應的古典測驗通過分數。

如此，再將每位考生的古典測驗原始答對題數總分大於或等於此通過分數者，即視為精熟者，低於此精熟分數者，則視為未精熟者。之後，再進行如上所

述的兩次交叉驗證分類，同時進行後續分類一致性的指標計算。

伍、轉換分數效益評估

對於轉換分數的效益評估方面，概念上可就幾方面探討之：首先，在評估古典測驗分數法與測驗特徵曲線構圖法於轉換 IRT- 能力值與古典測驗答對題數間之轉換效益前，得先釐清何謂轉換效益？何種條件下才可稱之為具有較佳的轉換效果？所謂轉換效益，在本研究中乃認為是研究者所期望的 IRT 精熟標準能力值與實際經轉換為古典測驗答對題數後，於運用測驗結果時的助益效果。

對此，加以詳細描述之，假設有以下兩種情況：在某次證照考試中，決策者乃決定採用 IRT- 能力值 0.5 為精熟標準，而經轉換後，相對之古典測驗分數為 60 分，此時，在進行受試者資格篩選時，研究者有兩方面選擇，第一，根據每位受試者 IRT- 能力值，以 0.5 作為能力篩選標準，而後，採用轉換後之古典測驗分數 60 分作為向大眾解釋之範例；第二，有別於依據受試者 IRT- 能力值進行能力篩選，在此，研究者顧及某些應用考量(如在發給考生的成績單中，必須註明其古典測驗總分時，考生可能在對照資格考試公告之古典測驗通過分數時，發現其分數已超過標準，但仍被判定為未精熟，此乃因以上述 IRT- 能力值作為內部篩選標準時，古典測驗分數 61 分對應 IRT- 能力值可能會小於 0.5，因而易造成解讀上的誤解)，乃以轉換後之古典測驗通過分數 60 分作為受試者篩選準則。此時，就上述第一種選擇而言，即無所謂轉換分數上分類的一致性，因為實際分類時，仍是依據受試者 IRT- 能力值進行篩選，而轉換之古典測驗答對題數則單純作為解釋用途，因而，對照於轉換分數效益上，只牽涉到所謂解釋上的效果(以古典測驗分數作為解釋時所收效益)；採第二種選擇時，轉換效益才會同時牽涉到解釋與轉換分數上分類一致性。

對照本研究設計，乃考量到上述的第二種選擇，探討採用換算古典測驗分數法與測驗特徵曲線構圖法於轉換分數上之精熟/未精熟者分類的一致性效果。轉換分數評估方面，乃強調著在同一份測驗下，由最大測驗訊息量法求得之 θ_{\max} 所作精熟/未精熟者分類，是否能與由換算古典測驗分數法或測驗特徵曲線構圖法轉換後之古典測驗通過分數所作精熟/未精熟者分類相一致，在此，輔以表 3-8 作範例說明，若由最大測驗訊息量法求得之 IRT 精熟標準 能力值與換算古典測驗分數法轉換之通過分數為 1 對 1 對應時，則在同一份測驗下(即隱含控制其它無關因素)，兩者分類精熟/未精熟人數應會相同(即 $A+B= A+C$ 或 $C+D= B+D$)，但若有轉換上的誤差時，則會出現如 B 與 C 的不一致精熟/未精熟分類人數。同理，在探討其它方法間的轉換效益時，亦是控制在同一份測驗下，檢視由方法求得之分類結果，而對照上述信度值即代表著高百分比一致性即具高的轉換分數效益。

表 3-8 轉換分數效益評估範例說明示意

第一次測驗		換算古典測驗分數法		合計
		精熟	未精熟	
最大測驗 訊息 量法	精熟	A	B	A+B
	未精熟	C	D	C+D
合計		A+C	B+D	

註：A 表由最大測驗訊息量法與換算古典測驗分數法同時分類為精熟者人數；B 表由最大測驗訊息量法分類為精熟者但換算古典測驗分數法卻視為未精熟人數；C 表最大測驗訊息量法分類為未精熟者但換算古典測驗分數法卻視為精熟人數；D 表最大測驗訊息量法與換算古典測驗分數法同時分類為未精熟者人數。

伍、定錨點

對於能力標準的描述，本研究乃沿用定錨點的概念，企圖描繪出精熟/未精熟者差異能力。對此，可從幾方面作深入探討，首先，在定錨點的選擇上，為期望能詳實描述精熟/未精熟者間差異能力，本研究乃援用 4 個定錨點，以精熟標準能力值為中心，選擇上下各 0.5 個標準差、1.5 個標準差能力值作為定錨點，相當於分別採用 2 個定錨點解釋精熟/未精熟者的能力；而於定錨點題目的選擇上，乃以各定錨能力值代入三個參數對數形模式之正確反應機率數學函數，以求得相對於每一試題之答對機率值，而後，採納 Huynh(1998)建議的三個準則：一、位於第 i 個定錨點上受試者，正確反應機率值大於或等於 .75；二、位於下一個定錨點受試者，正確反應機率，小於 .56(或為避免過於嚴苛，採取兩者中數 .66)；三、兩定錨點間試題正確反應機率差異大於或等於 .19。以此三準則作為篩選定錨點題目。最後，在定錨點描述上，主要以 Koretz & Deibert (1995)認為的以技能為基礎(skill based)的理念下，搭配國中基測能力指標，以描述精熟/未精熟受試者在內容領域上所具備差異知識與技能。而相關的測驗題本，可由國民中學學生基本學力測驗推動工作委員會網站上下載(網址：<http://www.bctest.ntnu.edu.tw/>)。

陸、精熟標準設定相關議題研究設計

本研究另針對精熟標準設定相關議題：測驗長度與測驗異質性，及其與上述方法間關聯進行探究，其設計方式茲簡述如下：

一、測驗長度

對於測驗長度議題設計，概念上期望能在控制測驗異質性、受試者能力分配因素下，探究不同測驗長度與精熟標準設定方法間可能的變化情形。實際執行時，乃針對自然科兩次測驗，採隨機方式各抽取出 10、20、30、40、50 題試題，

以分別代表不同的測驗長度類型，接續，利用試題產生樣本求得之已知試題參數，以計算兩次測驗研究操弄樣本 10000 名考生於各測驗類型下最大測驗訊息量法之 θ_{max} ，與換算古典測驗分數法、測驗特徵曲線構圖法之轉換的通過分數，進行如上述進行兩次精熟/未精熟者的交叉驗證分類分析與轉換效益探討，以期發現不同測驗長度類型與精熟標準設定間的互動關係。另外，與定錨點的關聯探討上，同樣是就不同測驗長度類型下，在固定 4 個定錨點時，以 Huynh(1998)建議之準則篩選定錨點題目，探討試題篩選可能受測驗長度類型影響的程度。

此外，對於 Huynh(1998)建議的第 2 個準則：位於下一個定錨點受試者，正確反應機率，小於 .56 或為避免過於嚴苛，採取兩者中數 .66 的兩種選擇上，對照於本研究篩選方式，則有些許調整，於正式解釋國中基測精熟/未精熟者差異能力時，是以嚴格 .56 標準，而於探討相關議題時，乃採用 .66 標準，此舉乃是企圖表達出在較寬鬆的情況下，至少能夠達到的某種定錨點題目篩選情況。

二、測驗異質性

對於測驗異質性設計，概念是希望探討在兩次測驗難度不同時，討論採用精熟標準設定方法於分類時受其影響的程度。執行時，乃將自然科兩次測驗，就試題產生樣本求得之已知試題參數，先依各自測驗試題難度值由低至高排序，而後，將較易的前 20 題與較難的後 20 題篩選出，即成為如圖 3-2 所示兩種測驗難度類型(難與易)。對此，再進一步說明，本研究篩選出各 20 題之難度類型試題，而非各 29 題的原因，其中除為了加大難與易測驗間的難度差異外，更是為了控制長度因素的干擾，方便直接援用測驗長度議題中 20 題之測驗以作比對，就此，乃將其稱之為常態測驗(有別於上述難與易測驗，其難度值較屬常態)。

接續，再針對各測驗難度類型，估計兩次測驗研究操弄樣本 10000 名考生最大測驗訊息量法之 θ_{max} ，與換算古典測驗分數法、測驗特徵曲線構圖法之轉換的通過分數，以便進行兩次精熟/未精熟者的交叉驗證分類分析。而於轉換效益評估上，有鑑於此設計概念乃強調著在「同一份測驗」下，由精熟標準設定方法與轉換通過分數方法間分類效果的一致性，概念上乃不同於測驗異質性的設計，強調著同一次分類比較中「兩次測驗」間難度的差異，因而，在此乃稍作調整，轉而探討不同分類比較間，採用簡易測驗、困難測驗或常態測驗間於轉換效益上是否會產生不同的效果？即是不同測驗難度類型是否會影響轉換效果？簡而言之，即是探討如第一次測驗 1-20 題(簡易測驗)求得之轉換效果，是否會明顯不同於第二次測驗 39-58 題(困難測驗)之轉換效果。

另外，與定錨點的關聯探討上，同樣是就不同測驗異質類型下，在固定 4 個定錨點時，以 Huynh(1998)建議之準則篩選定錨點題目，探討試題篩選可能受測驗異質類型影響的程度。

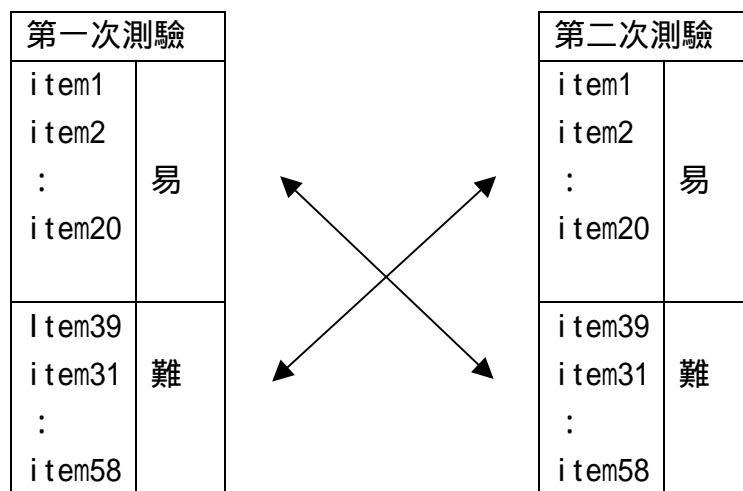


圖 3-2 測驗異質性研究設計圖

註：箭頭乃指交叉驗證之意。

第四節 資料分析與統計處理

以下茲就本研究所採資料分析方式，敘述如下：

- 壹、以 SPSS for Windows 11.5 版，進行樣本與試題選取、描述統計、因素分析、IRT- 能力值與古典測驗答對題數轉換、測驗特徵曲線答對機率值、選擇定錨點題目、百分比一致性及 κ 係數等統計數值的計算。
- 貳、採用 BILOG-MG 3.0 版的程式來估計受試者 IRT- 能力值、試題參數與最大測驗訊息量對應之 θ_{\max} 值的計算。其中，考量到受試者作答反應中會出現全對或全錯的情形，因此，根據 Hambleton & Swaminathan(1985)推導中，發現在最大概似估計法中，若受試者作答反應出現全對或全錯反應組型資料，該能力值只可能出現在正、負無限大時，因而，無法求得對照的能力值 (p.86)，而貝氏估計法則可經由預先確立的常態能力值的分配以解決此問題 (p.95)。有藉於此，本研究乃採行貝氏估計法(Bayesian estimation)中的期望後驗法(Expected A Posteriori, EAP)進行能力估計，而不採用其預設的能力估計法：最大概似估計法 (Maximum Likelihood Estimation)。