

## 國中學生英語閱讀電腦化測驗之編製與發展

廖麗珍<sup>1</sup> 邱毓芳<sup>2</sup> 郭伯臣<sup>3</sup> 白鎧誌<sup>3</sup> 白宗恩<sup>3</sup>

亞洲大學 資訊工程研究所<sup>1</sup>

國立臺中教育大學 教師教育研究中心<sup>2</sup>

國立臺中教育大學 教育測驗統計研究所<sup>3</sup>

lich@gm.cajh.chc.edu.tw, yvonnechiou65@gmail.com, kbc@mail.ntcu.edu.tw,  
minbai0926@gmail.com, smallwhite219@hotmail.com

### 摘要

本研究目的為依據國中英語教材和「九年一貫課程綱要」英語科中「讀的能力指標」，編製與研發仿教育會考的國中學生英語閱讀電腦化測驗，以彰化縣一所國中的345位三年級學生為樣本，檢驗其信度與效度。以試題反應理論分析，平均鑑別度為1.40，平均難度為0.066，平均猜測度為0.243，屬於難易適中的測驗，並且鑑別度良好；在測驗信度考驗方面，內部一致性信度 $\alpha$ 係數為0.896；效度採專家效度，請英語系教授審題，因此本測驗具良好信度與效度。此外，本研究以真實測驗資料模擬電腦化適性測驗施測程序，比較最大概似估計法、最大後驗估計法與期望後驗法在電腦化適性測驗評估的成效，研究結果顯示使用期望後驗法成效最佳。

**關鍵詞：**英語閱讀、電腦化測驗、電腦化適性測驗評估。

### Abstract

The purposes of this study were to develop an English reading computer-based test for junior high school students imitated from Education Examination based on junior English textbooks and English reading ability index for Nine-year Curriculum Outline. Three hundred forty-five ninth grade students sampled from an junior high school in Changhua participated in this study to verify the reliability and validity of the test, which was the main instrument in this study. Analyzing items with IRT, 3PL model, the mean of discrimination (a), difficulty (b), and guessing(c) are 1.40, 0.066 and 0.243 respectively. Therefore, it could be called a test with moderate difficulty and good discrimination percentage. According to the result of the reliability examination, the Cronbach value of internal consistency was 0.896. Moreover, criterion validity was based on expert validity. The researcher asked the English professors to revise the items. As a result, the English reading computer-based test for junior high school students has good reliability and validity. Besides, this study simulated computerized adaptive testing (CAT) measuring program with real

test data and compared the effects of computer adaptive testing assessments for MLE(maximum likelihood estimation), MAP(maximum a posteriori), and EAP (expected a posteriori). In the end, EAP turned out to be the most efficient in assessing the effects of CAT according to the findings of this study.

**Keywords:** English Reading, computer-based test, computer adaptive testing assessment.

### 1. 前言

隨著地球村時代的來臨，全球人類社會的互動更加頻繁，國際間政治、經濟、文化的交流愈為密切，被公認為世界共通語的英語，遂成為國際交流的重要溝通工具。此外，透過學習英文，學習者可以適切回應英語為母語的國家之社會文化活動並增進多元文化的瞭解與尊重。根據英國文化協會的報告，全世界約有八成以上的文獻資料是英文，而且大部分學者和科學家習慣以英文來閱讀。對許多學生來說，提昇英語閱讀能力和英語溝通能力成為適應當前競爭劇烈社會最迫切的需要。足見英文已是e世代吸收知識的關鍵能力，也讓政府意識到，英語能力之良窳，將決定下一代擷取知識的深度與廣度。在這樣的社會殷切期待下，自九十四學年度起英語教學已提早至國小三年級全面實施。

#### 1.1 研究動機

教學與評量並非兩條平行線，評量是教學當中不可或缺的重要部分，成功的教學，需透過評量的「程序」才能判定。教師能由評量的結果，瞭解教學的成果，以做為修正教學進度與掌握學生學習成效的參考，另一方面，具優良信效度的評量能提供學生適當的回饋，有助於瞭解自己的能力，藉此判斷需加強哪些單元的學習，足見評量之於教學的重要性。

國中教育會考(以下簡稱教育會考)是評量國中學生三年學習成果的重要測驗，教育會考的成績也是國中畢業生申請免試入學時作為超額比序的主要依據。而今年試辦的國中教育會考和國中基測一樣，受試學生仍需畫卡，然後再運用電腦讀卡與計

算成績，而此種測驗方式從出題、讀卡、到成績分析統計等，部分作業仍需靠人工完成。因此，隨著電腦科技的發達，愈來愈多的測驗，都改以電腦化測驗(computer-based test, CBT)進行，目前大型測驗如 GRE 和 TOEFL 已採行 CBT 或 IBT 模式施測。由於 CBT 和 IBT 具有不少優點，例如：(1)測驗的效率提高，能立即得知分數，快速處理相關的資料；(2)減少測驗時間，根據 Olsen、Maynes 和 Slawson [1]等人的研究指出，以電腦來實施同一版本的紙筆測驗，測驗時間只有紙筆測驗的 57%；(3)受試時間彈性大，受試者可隨時進行線上評量；此外，(4)可記錄解題歷程，CBT 能夠記錄受試者閱讀題目所需的時間、是否更改過答案、作答的順序等，讓教師瞭解學生的解題技巧以及困難所在，可幫助改進測驗或瞭解學生的障礙。

現行所有國中英語版本中，內容涵蓋很多重要的文法知識與艱深的句型，如關係代名詞、現在完成式、假設語氣…等較難理解的句型，使得多數學生覺得準備不易，對國三學生來說，教師教授新課程的時間已很緊湊，而教育會考時間約在國三下學期第二次段考後(5月中旬)，教師必須在此時間點之前教授完國三下學期課程並複習完國一、國二的課程實不容易！因此啟發研究者思考，若能編製一份符合教育會考英語科命題方向的試題並建置電腦化評量系統，讓學生可於課餘時間練習，又能得到立即回饋，有助於學生及時發掘自己的學習困難所在，亦利於教師實施補救教學。

## 1.2 研究目的

基於上述研究動機，研究目的如下：

- I 仿教育會考的命題架構編製「國中英語閱讀成就」測驗試題，檢驗其信度與效度。
- II 發展「國中英語閱讀學習成就」線上測驗系統。
- III 根據實徵資料分析最大概似估計法、最大後驗估計法與期望後驗法等三種能力估計法的電腦化適性測驗成效。

## 2. 文獻探討

本研究目的旨在依照「國民中小學九年一貫課程綱要」英語科第二階段能力指標編擬一份仿教育會考國中英語閱讀成就測驗的試題，對一群國三學生進行電腦化測驗，而電腦化測驗是以試題反應理論為基礎的測驗模式，故在本章分別就教育會考的試題編製、教育會考英語科的試題編製、試題反應理論及電腦化評量進行探討。

### 2.1 國中教育會考的試題編擬

教育當局為適度舒緩國中學生學習壓力，自

103 年起十二年國民基本教育政策將正式實施，中央及地方主管機關為了解及確保國中畢業生學力品質，將辦理「國中教育會考」，作為我國國中畢業生學力檢定之機制，而這項入學考試變革的宗旨即為在「降低壓力以活化學習」和「確保品質以維持競爭力」兩個目標之間求得平衡，換句話說，教育會考將在十二年國教的改革過程中，扮演適度降低壓力和維持競爭力的角色[2]。

#### 2.1.1 國中教育會考的試題編製

教育會考各學習領域試題之取材以學生學習及生活經驗為主要來源。每個題目都依明確的能力指標設計，並經過審查，以確保題目可以反映該能力指標。各學習領域一般命題原則如下：

1. 以能經由紙筆測驗評量的能力指標為主。
2. 以能有效檢測學生能力水準為目的。
3. 以符合綱要不涉及素材(版本)之選取為方針。

教育會考和國中基測一樣，試題係依據九年一貫課程綱要國中階段能力指標來命題，採教材(內容)、能力之雙向評量方式研發，測驗的精神在評量學生基本、核心、重要的知識與能力，但測驗試題難度以「難易適中」為原則[3]。

教育會考為一標準化測驗，依「闡外建置題庫，闡內依組題計畫抽組試題」為試題研發方式，茲說明組題計畫如下：

教育會考的題本由闡場的組題電腦系統來選題，無法事先得知系統組成的測驗試題，其組題的原則為評量需符合教育會考目的，目的為能精確區分「基礎」與「精熟」能力等級的學生；試題取材均勻分布；生活化、統整化素材優先；試題平均通過率為 .5~.6，另外，林世華指出試題難度低於30%的不可出現，因為通過率低於30%時，就不算是基本的試題[4]。

題庫中的測驗組合系統可使選題程序電腦化，以避免人工選題產生的疏失與不公正的現象。在選題策略上，國中教育會考使用 0-1 線性規劃來作為選題程式的主幹，依據題目的各種屬性(例如：題目的類型、認知層次、所屬的章節範圍…等)，以及各項統計上的指標(例如：題目的難度、鑑別度、訊息量…等)，設計出線性方程組的限制條件，再配合線性規劃求取目標函數的極大(小)值進行組題。相較於隨機選題或是人工組題，經由這樣的組題程序所獲得的題本，在難度及類型的題數分布上，較為客觀[5]。

#### 2.1.2 國中教育會考英語科試題的編製

教育會考英語科試題的研發係以「國民中小學九年一貫課程綱要」英語科國中階段能力指標為依據，試題編寫以一千兩百基本字彙為主，考生在一

般學習狀況下習得基本英語聆聽及閱讀能力，不論使用哪一版本，均能夠作答，不須習讀所有版本。

依據課程綱要，國民中小學英語課程目標為：

1. 培養學生基本的英語溝通能力，俾能運用於實際情境中。
2. 培養學生學習英語的興趣與方法，俾能自發有效地學習。
3. 增進學生對本國與外國文化習俗的認識，俾能加以比較，並尊重文化差異。[6]

在課程綱要中，英語科課程分為國小和國中兩個階段，兩階段之教學重點是不同的。國小以聽、說為主，讀、寫為輔；國中則聽、說、讀、寫並重。

有關英語國中階段讀的能力指標為：

- 3-2-1 能辨識英文字母的連續書寫體。
- 3-2-2 能用字典查閱字詞的發音及意義。
- 3-2-3 能看懂常用的英文標示和圖表。
- 3-2-4 能用適切的語調、節奏朗讀短文、簡易故事等。
- 3-2-5 能瞭解課文的主旨大意。
- 3-2-6 能瞭解對話、短文、書信、故事及短劇等的重要內容與情節。
- 3-2-7 能從圖畫、圖示或上下文，猜測字義或推論文意。
- 3-2-8 能辨識故事的要素，如背景、人物、事件和結局。
- 3-2-9 能閱讀不同體裁、不同主題的簡易文章[6]。

國中教育會考英語科閱讀測驗主要是評量基本重要概念，非片斷的記憶，試題編寫重視真實語言使用的整體自然情境，呈現多種體裁及主題。

國中教育會考英語科聽力與閱讀測驗合計共60題，其中前20題為聽力試題，後40題為閱讀試題。本研究範疇為閱讀學習成就測驗試題。題本架構包括單題和題組兩種題型，其中單題部分主要在評量語言知識，如字彙、語意和語法；題組則是評量篇章理解能力，包括克漏字和整段式閱讀測驗文章。以102年試辦國中教育會考英語科閱讀測驗的題組選文而言，除了文章及對話等一般體裁，也有電子郵件、圖表、詩歌、網頁廣告及新聞報導等各式體裁[7]。

## 2.2 試題反應理論

為了改進古典測驗理論的缺失，因而產生當代測驗理論[8]，當代測驗理論以試題反應理論為基礎，是一具有合理與嚴謹的立論與假設的學說，採用複雜艱澀的計算公式，所適用的測驗資料種類雖屬有限，但深受測驗學者的青睞，有後來居上之勢。

試題反應理論(item response theory, IRT)又稱作潛在特質理論，建立在兩個基本概念上：(1)受試者的答對機率受到本身潛在特質或能力的影響[9]。(2)受試者的表現與潛在特質間的關係可由一條連續性的試題特徵曲線來加以詮釋[10]。

試題反應理論為試題獨立的能力估計值，不會受不同測驗種類的影響；其次，它是樣本獨立的試題參數估計值，不會受到不同族群考生的影響。

IRT有四項基本假設：單向度、局部獨立性、非速度測驗與知道—正確假設[11]。而IRT的試題參數模式可分為單參數模式(Rasch模式)、二參數模式和三參數模式[12]，都只適用於二元化的反應資料。教育會考和基測一樣，測驗題庫均採用Rasch模式，在此模式中影響受試者答對率的試題參數只考慮難度參數，未考慮鑑別度與猜測度等參數，較無法呈現試題的內部一致性信度，故本研究將採三參數模式做試題分析。

三參數模式(three-parameter logistic model)可用公式1表示：

$$P_i(\theta) = C_i + (1 - C_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \quad i = 1, 2, \dots, n \quad (1)$$

其中， $P_i(\theta)$ 表示能力為 $\theta$ 的考生在試題 $i$ 上正確反應的機率； $b_i$ 表示試題難度(difficulty)參數； $a_i$ 表示試題鑑別度(item discrimination)參數； $c_i$ 表示猜測度參數(pseudo-chance parameter)； $n$ 是該測驗的試題總數； $e$ 代表以底為2.718的指數；且 $P_i(\theta)$ 是一種S形曲線，其值介於0與1之間[12]。

## 2.3 電腦化評量系統

電腦化測驗已是世界發展的趨勢，除了可運用多媒體科技，讓試題以色彩、聲音、動畫、互動、操作等方式呈現，測驗資料可進行試題分析及信度與效度估計，此外，近年來由於網路頻寬加大及資料傳輸速度提高，透過網路實施電腦化適性測驗也漸漸增加。目前電腦化測驗系統的建置多以IRT模式做為理論基礎，較複雜的模式也引進到電腦化測驗系統中，如：CAT (Computer Adaptive Testing)。目前國外已經發展電腦化評量系統的如美國軍方電腦化職業性向測驗系統、英國劍橋大學英文考試院(Cambridge ESOL)發展出的BULATS英語能力檢定和ETS(Educational Testing Service)所承辦的TOEFL、GRE等測驗；國內則包含TOEFL測驗、交通部公路局的汽機車駕照考試和考選部的航海類專技人員考試都已實施電腦化測驗。

## 3. 研究方法

本研究旨在編製仿教育會考國中英語閱讀成就測驗試題和發展線上測驗系統。根據研究動機、研究目的，透過文獻與國中現行教科書其中四個版本的分析與探討，並參考教育會考的題本架構，編製適合國三學生的閱讀學習成就測驗試題，並建置電腦化測驗系統以作為學生自我評量的工具。

### 3.1 電腦化測驗研發

本研究蒐集近兩年的模擬考題本和歷年基測試題，諮詢領域專家與國中現職英語教師之意見，決定以 1,200 基本字彙為可命題的字彙，接著根據教育會考英語科的命題原則和國中階段讀的能力指標之檢核重點，從模擬考題本和基測試題中挑選基本、核心概念和重要知識的試題，與國中現職英語教師就題目內容進行討論，並請英語學者幫忙修審試題，以確保試題品質和可行性。表 1 為正式組卷的試題分布情形如下：

表 1 試題分布

能力指標	分佈題號
3-2-1	第 2, 3 題
3-2-3	第 27, 28, 29, 30 題
3-2-5	第 24, 26, 39 題
3-2-6	第 4, 5, 8, 11, 13, 14, 16, 19, 25 題
3-2-7	第 1, 6, 7, 9, 10, 12, 15, 17, 18, 31, 32, 35 題
3-2-8	第 20, 21, 22, 23, 33, 36 題
3-2-9	第 34, 37, 38, 39 題

其中能力指標 3-2-2 和 3-2-4 實務上無法透過電腦化評量系統施測，需耗費很多人力當面施測，是故，不適合研發相關題目。

組卷完成後，本研究直接開發「電腦化測驗系統」，提供使用者由 IE 瀏覽器進入系統施測。以下為系統介面步驟說明和執行畫面：

#### 一、登入系統

每位受試者會有自己的帳號、密碼登入系統施測，如圖 3.1 所示。



圖 3.1 登入系統畫面

#### 二、進行電腦化測驗，如圖 3.2 所示。



圖 3.2 進行電腦化測驗畫面

### 3.2 測驗資料分析

為瞭解試題品質，研究者採用測驗的實徵資料進行試題分析，在估計過程中無法收斂者、試題通過率偏低等情況的試題予以刪除，重新估計後所餘試題為 39 題。

為了顧及測驗資料品質，本研究參考國外大型測驗的檢核程序，訂定仿教育會考國中英語閱讀成就測驗的無效樣本檢核準則，簡述如下：

- (1) 考生的作答反應有疑似亂答情況(規則性作答)
- (2) 考生整份測驗作答時間在 30 秒內，且答對率低於 0.25

關於不良試題的檢核準則，參考 IRT 估計之參數值，訂定標準如下：

1. IRT 的試題鑑別度參數小於 0.4 ( $a < 0.4$ )
2. IRT 的試題難度參數小於 -2 ( $b < -2$ )；或者大於 2 ( $2 < b$ ) [12]

本研究使用 SPSS 進行內部一致性信度分析，藉著信度估計來檢定測驗中試題之一致性或關聯性，若 Cronbach's  $\alpha$  值很低，表示試題之間的關連性很低，也就是測驗不具信度。針對一般性的研究來說，內部一致性估計可接受數值為 .80，標準化成就測驗  $\alpha$  係數最好在 .9 以上；至於效度分析方面，本研究透過專家學者對於內容上的專業知識，進行題項與形式之審查，是故，本研究是採專家效度。

本研究以 BILOG-MG 軟體進行 IRT 各模式之試題參數估計，由於本測驗採選擇題的方式，含猜測成分，因此採三參數的估計，即鑑別度、難度和猜測度參數。此外，利用 IRT 來分析本研究之測驗資料，需滿足 IRT「局部獨立」的假定。本測驗資料分析包括 39 道試題，每道題目均測量不同的語意或句型，亦即每一題的答對機率不會因受試者在其他題目的表現所影響。

### 3.3 電腦化適性測驗評估

電腦化適性測驗 (Computer Adaptive Testing, CAT) 的原則是依照受試者的作答反應估計出其能力值，然後提供適合受試者能力的題目進行施測，然而在測驗開始時，並不知受試者能力值，所以必須先決定測驗的起始點，以選擇試題供受試者施測。常用的選題法如：依受試者能力選題、中等難度題目選題、自由選題、隨機選題等。本研究目的為建置一個完整的電腦化測驗系統 (computer-based test, CBT)，不過目前電腦化適性測驗已是教學評量的發展趨勢，啟發研究者思考，使用測驗的實徵資料進行電腦化適性測驗評估，而為了進行適性測驗評估，採用較常見的能力估計方法，例如：最大似估計法 (maximum likelihood estimation, MLE) 與貝氏估計法 (Bayesian procedure) 中的最大後驗估計法 (maximum a posteriori, MAP) 與期望後驗法

(expected a posteriori, EAP), 以下分別介紹：

a. 最大概似估計法 (MLE)

假設測驗共有  $n$  道試題，試題間彼此獨立，概似函數 (likelihood function) 如下式 (2) 所示：

$$L(u|\theta) = L(X_1, \dots, X_n|\theta) = \prod_{i=1}^n P_i^{X_i} Q_i^{1-X_i} \quad (2)$$

◆為所有作答反應的向量， $L(X_1, \dots, X_n|\theta)$  為概似函數； $\theta$  為受試者的真實能力； $X_i$  指受試者在第  $i$  題的作答反應，答對為 1，答錯為 0； $P_i$  指受試者在第  $i$  題的答對機率； $Q_i$  指受試者在第  $i$  題的答錯機率 [13]。

b. 期望後驗估計法 (EAP) [14]

Bock 與 Mislevy 主張 EAP 是尋找能力值的事後機率密度函數的期望值，如公式 (3) 所示：

$$\theta_{EAP} = \frac{\sum_{q=1}^{k_q} \theta_q f(\theta_q|U)}{\sum_{q=1}^{k_q} [L(U|\theta_q) f(\theta_q)]} \quad (3)$$

$u$  為所有作答反應的向量， $L(U|\theta_q)$  為概似函數； $\theta_q$  為受試者的真實能力； $q$  是計算能力的期望值時所切割成的分割點，共有  $k_q$  點， $k_q$  愈大，計算愈精確。

c. 最大後驗估計法

MAP 是以受試者的事前能力分布  $f(\theta)$  作為加權值，形成事後機率密度函數，並找出能使此事後機率密度函數最大化的程度值。事後機率密度函數的計算式 (4) 如下：

$$f(\theta|U) = \frac{L(U|\theta) f(\theta)}{f(U)} \quad (4)$$

$L(U|\theta)$  是受試者  $\theta$  的概似函數， $f(U)$  是受試者的實際機率，是由  $L(U|\theta) f(\theta)$  從  $-\infty \sim \infty$  積分所得，若要加速找到事後機率密度函數的最大值，通常以 Newton-Raphson 法來進行迭代。

本研究以上述之能力估計法進行參數估計值誤差評估以診斷出能最快達到適性目的的能力估計法；而參數估計值誤差評估的過程為視原始生成參數為真值，以真值產生作答反應，再以作答反應進行參數估計後得到估計值，計算真值與估計值的均方根誤差 (root mean squared error, RMSE)。當均方根誤差值在 0.2 以下時，測驗即可終止，因為此種情形就是適性測驗。以上三種能力估計法估 RMSE 是根據能力值計算。其數學式為：

$$RMSE(\theta) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\theta_i - \hat{\theta}_i)^2}$$

$\hat{\theta}_i$ ：第  $i$  位能力值估計值

$\theta_i$ ：第  $i$  位能力值真值

$N$ ：受試者人數

### 3.4 研究對象

本研究以彰化縣一所國中三年級 14 個班級的學生為研究對象，共計 410 位，其中有效樣本為 345

人。

### 3.5 研究工具

a. 基測和教育會考試題

本研究利用四個版本國中英語教科書的基本、核心、重要的教材內容來檢核試題，並配合九年一貫國中英語科讀的能力指標以分析基測試題及試辦教育會考考題的命題原則，以上述分析的結果作為依據，評估本研究編製的試題內容、認知層次是否符合國中教育會考的方向，然後形成線上測驗的正式組卷。

b. 線上測驗系統

研究運用的線上測驗系統是由國立臺中教育大學教育測統所研發的「電腦化測驗診斷系統」，針對研究對象進行線上電腦化測驗與成效評估。

c. 各項統計分析軟體

BILOG-MG 軟體

本研究使用 BILOG-MG 軟體能夠估計測驗試題的參數值，包括難度、鑑別度、猜測度及受試者的能力值 [15]。

SPSS 統計分析軟體

統計分析軟體 SPSS 是用於大量資料的統計分析，本研究以 SPSS 分析測驗的信度及相關係數。

MATLAB 軟體

利用 BILOG-MG 估計的試題參數值和受試者能力值，再透過 MATLAB 軟體撰寫 MLE、EAP、和 MAP 等三種能力估計方式，以分析出何種方法可使本研究的電腦化測驗最快達到適性的效果。

## 4. 研究結果

依據前述之研究方法，本章以研究對象施測之資料為基礎，透過測驗理論和研究工具針對測驗結果進行分析和估計，此外，研究者利用不同能力估計方法評估電腦化適性測驗成效。以下為研究結果的分析與討論。

### 4.1 測驗結果描述

本研究測驗信度 Cronbach's  $\alpha$  值為 .896，顯示測驗信度佳；而效度檢定採專家效度，研究者邀請兩位國立臺中教育大學英語系教授依據自身在學科上之專業知識，進行試題內容與題型之審查。

實施電腦化測驗後，依自訂的無效樣本指標，刪除 65 位無效樣本，共 345 位有效樣本；並依據 IRT 的不良試題檢核標準，刪除試題，剩餘試題共 39 題。接著，以 BILOG-MG 軟體做 IRT 之試題參數估計，顯示平均難度參數為 0.066，符合教育會考「難易適中」原則；試題平均通過率為 0.601，大致符合教育會考試題平均通過率 0.5~0.6 的標準；平均鑑別度參數為 1.395，顯示測驗具良好鑑別度；平均猜測度值為 0.243，符合 IRT 猜測度參



數小於 0.3 的指標，而受試者的能力估計值為 -2.248~2.215 之間，在正常值 -4~+4 範圍。

#### 4.2 電腦化適性測驗評估結果

本研究以真實測驗資料模擬 CAT 施測程序，比較最大概似估計法 (maximum likelihood estimation, MLE)、最大後驗估計法 (maximum a posteriori, MAP) 與期望後驗法 (expected a posteriori, EAP) 不同能力估計法在電腦化適性測驗評估的成效，三種能力估計法模擬 CAT 施測成效如圖 4-1。

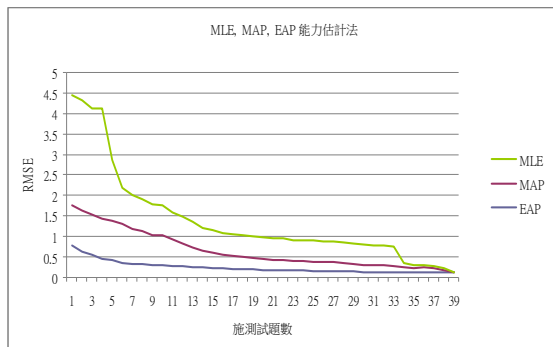


圖 4-1 不同能力估計方法之精確度

圖 4-1 為使用三種能力估計方法在不同測驗長度下與施測全部試題 RMSE 的變動情形，圖 4-1 顯示不論使用哪一種估計方法，當施測題數累積愈多，RMSE 的下降就愈明顯。從估計結果顯示 MLE 估計法測驗題數在 30 題之前，RMSE 都大於 0.5；總題數達到 34 題時，RMSE 小於 0.2；MAP 估計法施測總題數在 12 題以前，RMSE 都大於 0.5；當總題數達 29 題時，RMSE 小於 0.2。EAP 估計法施測總題數達 4 題時，RMSE 小於 0.5；總題數達 19 題時，RMSE 小於 0.2。在這三種估計法中，EAP 估計法得到較低的 RMSE，此結果與陳柏熹提出的研究結果類似，因此在適性功能的評估中，使用 EAP 估計法成效較佳〔16〕。

#### 5. 結論與未來研究方向

透過研究樣本測驗資料分析結果顯示，本研究研發之測驗信度介於 0.890~0.903 間，且具有良好的效度，整份試題都經過領域學者專家縝密修審，具有一定的品質，未來可做為一般國三學生英語閱讀能力測驗檢測之工具。

本研究使用 IRT 三參數模式能改善 CTT 模式的諸多缺點，且和教育會考測驗題庫僅採用 Rasch 模式分析測驗資料比較，本研究較能呈現內部一致性信度。

本研究研發之仿教育會考英語閱讀成就測驗配合英文電腦化評量系統進行施測，教師可透過系統介面管理者得知學生的施測成績，作為補救教學的依據；使用者可隨時登入系統介面，以帳號和密

碼註冊回傳即進行線上測驗，施測完可立即看測驗結果報告得到回饋；管理者能夠在試題庫管理介面隨時新增或修改試題，也可取得受試者作答反應。

未來將視教科書內容變動，更新或增修試題，也計劃擴充試題規模建立題庫，以國立臺中教育大學教育測統所建構的適性化評量系統為平台，使用者由 IE 瀏覽器進入平台並註冊帳號、密碼，即可開始測驗。經由此適性評量系統，使用者可依個別能力差異進行不同長度的測驗，達到省題和省時的目標。

#### 參考文獻

- [1] Olsen, J. B., Maynes, D. D., & Slawson, D., Comparison and equating of paper-administered, computer-administered, and computerized adaptive tests of achievement, San Francisco: Paper presented to the annual meeting of the American Educational Research Association, 1986.
- [2] 教育部，十二年國民基本教育落實國中教學正常化、適性輔導及品質提升方案，中央及地方主管機關，2012。
- [3] 臺灣師大心測中心，國中教育會考試題取材與命題原則，（台北市，2012）。
- [4] 林世華，國民中學學生基本學力要如何走下去，（飛揚月刊，第九期，2008），取自 <http://www.bctest.ntnu.edu.tw>
- [5] 宋曜廷、許福元、曾分蘭、蔣莉蘋、孫維民，國民中學學生基本學力測驗的回顧與展望，（台北：教育研究與發展期刊，第三卷第四期，2007），pp. 34-38。
- [6] 教育部，國民中小學九年一貫課程綱要語文學習領域（英語），（台北市，2008）。
- [7] 教育部，國中教育會考英語科加考聽力測驗說明，台北市，2012。
- [8] 王寶瑯，現代測驗理論，臺北：心理出版社，1995。
- [9] Hambleton, R. K., & Cook, L. L., Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement*, 14, 1977, 75-96.
- [10] Lord, F. M., Practical applications of item characteristic curve theory. *Journal of Educational Measurement*, 14, 1977, pp. 117-138.
- [11] 余民寧，試題反應理論的介紹（二）— 基本概念和假設，*研習資訊* 9 卷 1 期，1993，pp. 5-9。
- [12] 余民寧，試題反應理論的介紹（三）— 試題反應模式及其特性，*研習資訊* 9 卷 2 期，1993，pp. 6-10。
- [13] Bock, R. D., & Aitkin, M., Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 1981, pp. 443-459.
- [14] Bock, R. D. & Mislevy, R. J., Adaptive EAP Estimation of Ability in A Microcomputer Environment, *Applied Psychological Measurement*, 6, 1982, p.431-444.
- [15] Birnbaum, A., Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, *Statistical theories of mental test scores*, 17-20. Reading, MA: Addison-Wesley, 1968.
- [16] 陳柏熹，能力估計方法對多向度電腦化適性測驗測量精確度的影響，*教育心理學報*，38 (2), 2006, pp.195-211.