

華語文能力測驗技術報告 — 2013(2)

閱讀測驗信效度

國家華語測驗推動工作委員會 編著

序

中文近年來已經成為許多國家優先學習的第二外國語言，中文教學（又稱華語教學）也成為教育界新興的重要學科。世界各國的華語學習者越來越多，開始學習華語的年齡也持續下降，充分顯示華語的國際影響力與華語教學的未來發展潛力。

良好的華語教學除了要有優良的師資外，還需要仰賴優質的課程設計、教材及測驗工具。臺灣師範大學於 1956 年開始投入華語教學，這五十年來已經累積相當豐富的經驗，成為全球華語教學重鎮。我們也在教育部委託下，於 2005 年成立國家華語測驗推動工作委員會，並積極發展各式華語能力測驗，希望建立具有公信力的標準化華語測驗，成為全球知名的華語能力評量工具。

為了能有效評量出學習者的不同華語文能力水準，並且與國際的語言能力學習暨評量架構接軌，本會特別研發了不同等級的聽力、口語、閱讀以及寫作「華語能力測驗」，以及針對兒童所研發的「兒童華語文能力測驗」。測驗內容參考了歐洲共同語文參考架構，以及學習者的學習時數與華語能力發展，題材涵蓋食衣住行…等多元面向。為了讓學習華語者及華語相關領域工作者能更明白本會發展之各測驗的內容、測驗實施方式與成績公佈之標準化流程、以及本年度正式考試之測驗效能評估，我們特別出版這份技術報告供大眾參考。

國家華語測驗推動工作委員會執行長 陳柏熹 謹識

目錄

一、	前言.....	1
二、	簡介.....	2
	(一) 能力描述.....	2
	(二) 測驗題型與題數.....	2
	(三) 通過門檻.....	4
三、	測驗標準化流程.....	7
	(一) 試題收集.....	7
	(二) 試題修審.....	8
	(三) 預試.....	8
	(四) 試題分析.....	9
	(五) 輸入題庫.....	9
	(六) 組成正式卷.....	9
	(七) 檢核正式卷與多媒體檔案.....	10
四、	測驗評估.....	11
	(一) 信度.....	11
	1. 內部一致性信度.....	12
	2. 測驗訊息量.....	12
	(二) 效度.....	15
	1. 內容效度.....	16
	2. 建構效度.....	17
五、	結論.....	36
六、	文獻.....	37

表目錄

表 1	基本能力描述	2
表 2	測驗題型與題數分布	3
表 3	標準設定各回合判斷結果之標準差	5
表 4	測驗通過分數	6
表 5	各等級正式考試卷難度分布	10
表 6	內部一致性信度摘要表	12
表 7	雙向細目表	17
表 8	試題難度估計分布	18
表 9	受測者能力估計分布	18
表 10	試題適配分布	21
表 11	入門基礎級測驗不適配試題選項描述性摘要表.....	22
表 12	進階高階級測驗不適配試題選項描述性摘要表	27
表 13	流利級測驗不適配試題選項描述性摘要表	29
表 14	整體模式適配度指標摘要表	35

圖目錄

圖 1	正式考試標準化流程	7
圖 2	入門基礎級測驗訊息量	14
圖 3	入門基礎級測驗測量誤差	14
圖 4	進階高階級測驗訊息量	14
圖 5	進階高階級測驗測量誤差	15
圖 6	流利級測驗訊息量	15
圖 7	流利級測驗測量誤差	15
圖 8	入門基礎級測驗受測者與試題分布圖	19
圖 9	進階高階級測驗受測者與試題分布圖	19
圖 10	流利級測驗受測者與試題難度圖	20
圖 11	RIA_04 試題特徵曲線	23
圖 12	RIA_05 試題特徵曲線	23
圖 13	RIA_21 試題特徵曲線	24
圖 14	RIA_41 試題特徵曲線	24
圖 15	RIIA_13 試題特徵曲線	25
圖 16	RIIA_21 試題特徵曲線	25
圖 17	RIIA_39 試題特徵曲線	26
圖 18	RIB_46 試題特徵曲線	28
圖 19	RIB_50 試題特徵曲線	28
圖 20	RIIB_14 試題特徵曲線	29
圖 21	入門基礎級閱讀測驗三因素驗證性因素分析	32
圖 22	進階高階級閱讀測驗四因素驗證性因素分析	33
圖 23	流利級閱讀測驗二因素驗證性因素分析	34

附件目錄

附件 1	閱讀測驗各等級標準設定會議問卷調查結果.....	39
附件 2	入門基礎級測驗各題因素負荷量及殘差變異量摘要表.....	41
附件 3	進階高階級測驗各題因素負荷量及殘差變異量摘要表.....	44
附件 4	流利級測驗各題因素負荷量及殘差變異量摘要表.....	47

一、 前言

「華語文閱讀測驗」(以下簡稱本測驗)是專為母語非華語者所研發的一套標準化語言能力測驗。「國家華語測驗推動工作委員會」(以下簡稱本會)自 2001 年 8 月開始籌劃本測驗，2003 年 12 月正式對外開辦考試。閱讀測驗旨在測知華語學習者於實際日常生活中的語言使用能力，故不以任何特定教材為命題依據。測驗內容主要針對各種日常生活情境所設計，題材真實多元，涵蓋多種不同體裁，如廣告、公告、信件、便條、天氣預報等。所選短文之主題也均衡取材自文學、科學、藝術、天文、歷史等不同領域。為使本測驗與歐洲共同語文參考架構(Common European Framework of Reference for Languages: Learning, Teaching, Assessment; 簡稱 CEFR)接軌，提供考生衡量其語言能力的評量工具，本會於 2008 年起著手進行測驗的改版工作¹。

自 2013 年起，「華語文閱讀測驗」的測驗架構分成三等六級，三等分別為入門基礎級、進階高階級及流利精通級，而每一等又可再依據測驗成績細分為兩級，分別為入門級、基礎級、進階級、高階級、流利級、精通級²，共六級。應試者可依自己的學習背景或語言能力選擇合適的等級應考。此架構相較於僅能區分受測者是否通過測驗而言，能夠更進一步區分出通過測驗的受測群體其能力的高低；同時，對於應試者及試務工作者來說，更符合經濟效益。例如，改版後的測驗方式(一等兩級)，應試者可依自己的學習背景或語言能力選擇合適的等級應考，讓考生只要參加一次測驗，即可同時判斷兩等級程度。

本報告首先針對 2013 年華語文閱讀測驗能力描述、題型、等級通過門檻，以及測驗的實施與成績公布之標準化流程進行概述。其次，分別闡述該年度正式考試之信效度分析結果。最後，根據各項分析結果提出相關討論及建議。

¹ 各等級閱讀測驗分別對應於歐洲理事會(2001)所頒布的「歐洲共同語文參考架構」(Common European Framework of Reference for Languages: Learning, Teaching, Assessment; 簡稱 CEFR)之各能力等級。

² 2013 年本會發展精通級預試，預計於 2014 年推出流利精通級正式考試。

二、 簡介

2013 年度本測驗正式考試等級分為：入門基礎級(Band A)，區分為入門級(Level 1)與基礎級(Level 2)；進階高階級(Band B)，區分為進階級(Level 3)與高階級(Level 4)；及流利級(Level 5)。並分別對應於歐洲共同語文參考架構(CEFR)之 A1(Breakthrough)、A2(Waystage)、B1(Threshold)、B2(Vantage) 與 C1(Effective Operational Proficiency)。以下分別就各等級能力描述、測驗題型與題數、通過門檻三個方面進行介紹。

(一) 能力描述

各等級中，測驗通過者所需具備的基本閱讀能力如表 1 所示。其中，入門基礎級側重於「日常生活中簡易句子的理解能力」；進階高階級以「語言段落的理解分析能力」為主；流利級則著重於「語言使用的廣度與精熟度」。

表 1 基本能力描述

測驗等級	能力描述
入門級	在有視覺協助及可重複閱讀的情況下，能掌握基本數字、詞彙及簡單的短語並能大致理解句子內容。
基礎級	能理解用日常生活詞彙或工作常用詞彙寫成的簡短文章。
進階級	能讀懂個人感興趣的主題或與專攻領域相關的文章；前提是文章以淺白、平鋪直敘的方式寫作而成。
高階級	閱讀具有相當大的自主性，懂得為了不同目的，採用不同方法和速度閱讀不同的文章，並能選擇適合使用的參考書。具備廣泛且可隨時提取的閱讀詞彙，但對於不常見的慣用語，可能有理解上的困難。
流利級	在有機會重新閱讀困難部分的情況下，不論主題是否與個人專攻領域相關，都能讀懂長篇複雜文本的各項細節。

(二) 測驗題型與題數

本測驗根據各等級語言學習者的能力表現(如表 1)以及測知目標的差異，在入門基礎級測驗共採用「單句理解」、「看圖釋義」、「選詞填空」以及「完成段落」等四個題型；進階高階級與流利級則採用「選詞填空」與「短文閱讀」等兩個題

型，各等級題型與題數如表 2 所示。入門基礎級前四部分為三選一之單擇題，第四部分為配合題；進階高階級與流利級兩部分皆為四選一之單選題。各測驗等級總題數為 50 題，測驗時間為 60 分鐘。

表 2 測驗題型與題數分布

測驗等級	題型	題數
入門基礎級	單句理解	20
	看圖釋義	15
	選詞填空	10
	完成段落	5
進階高階級	選詞填空	15
	短文閱讀	35
流利級	選詞填空	15
	短文閱讀	35

由表 2 可發現，各等級皆包含「選詞填空」題型，其原因為「選詞填空」主要是測得語言知識的題型，而在外語的學習過程中，語言知識是建立語言能力的基礎(Bachman, 1985; Cohen, 1994; Oller, 1979)，因此，「選詞填空」題型在各等級測驗中均被採用。本測驗之「選詞填空」題型是採用了多項選擇式的克漏字測驗(Multiple-choice Cloze Test)搭配意向刪詞法(rational deletion)設計而成，目的是為了達到設計出不同體裁、難度的文本(text)以助於有效評量各等級學習者能力(洪小雯，2009)。

此外，當評量學習者的閱讀表現時，語言的複雜度、文本類型、篇章結構與長度等特性，都應列入考量。例如，語法結構過於複雜會消耗受測者對測驗內容的注意力；具體的描述與指示(特別是有足夠的視覺輔助時)可能比抽象的辯論或解釋來得容易理解(Council of Europe, 2001, 頁 165)。由於入門基礎級受測者的語言能力尚處於初級階段，為減少其因處理文字的負荷量過大而產生焦慮，進而影響作答，因此，在入門基礎級測驗中多採用情境圖片輔助試題文字；相較於入門基礎級受測者，進階高階級受測者的語言能力較為純熟，故在進階高階級測驗中圖文相輔的題型(如材料閱讀)較少；而流利級的語言能力已不需要額外輔助即可了解文本內容(Council of Europe, 2001)，所以，在流利級測驗中皆未設計圖文輔助題型。

(三) 通過門檻

本測驗共透過三次標準設定(standard setting)程序，以 Yes/No Angoff 法 (Impara & Plake, 1997)設定出入門級、基礎級、進階級、高階級、流利級與精通級六個等級之通過門檻(藍珮君等，2013；Chen et al., 2014)。所有標準設定成員均由華語文及語言學領域專家所組成，各次標準設定程序皆依循相同標準化流程執行。以設定入門基礎級測驗門檻為例，標準設定程序各步驟說明如下，詳細內容可參考藍珮君等人(2013)與 Chen 等人(2014)之研究。

1. 簡介此標準設定之目的與門檻設定的方法(Yes/No Angoff 法)，並介紹試題反應理論(item response theory；簡稱 IRT)模式中的 Rasch 模式(Rasch, 1960)(如公式 1)。

$$P(X_{ni}=1|\theta_n,\delta_i)=\frac{\exp(\theta_n-\delta_i)}{1+\exp(\theta_n-\delta_i)} \quad (1)$$

其中， θ_n 表示受測者 n 的能力； δ_i 表示試題 i 的難度(difficulty)； $P(X_{ni}=1|\theta_n,\delta_i)$ 表示受測者 n 在試題 i 的答對機率。介紹 Rasch 模式的原因為本測驗試題發展採用之理論模式為 Rasch 模式。

2. 介紹 CEFR 架構，並說明本測驗依據 CEFR 之 A1 及 A2 等級能力描述所定義之入門級與基礎級最低能力描述(minimum performance level descriptions)。
3. 門檻設定成員於一份 50 道閱讀測驗試題之判斷表中開始第一回合門檻設定判斷。
4. 提供成員根據步驟 3 的判斷結果所得之回饋訊息(Cizek & Bunch, 2007)。回饋訊息包含：(1)透過判斷結果與試題難度參數，得出入門級與基礎級通過門檻的答對題數與受測者能力值的平均數及標準差；(2)50 道試題之難度參數；及(3)以暫時門檻得出之受測者通過入門級與基礎級之比例。接著，成員們依據上述回饋訊息進行第一回合判斷後討論。
5. 完成第一回合討論後，成員再次從 50 道試題判斷表中進行第二回合門檻設定判斷，判斷方式同步驟 3。
6. 根據步驟 5 之第二回合判斷結果，提供成員如步驟 4 之回饋訊息，並進行第二回合判斷後討論。

7. 完成第二回合討論後，成員再次從 50 道試題判斷表中進行第三回合門檻設定判斷，判斷方式同步驟 3 及 5。
8. 依據成員於步驟 7 所設定之門檻及本測驗發展目的與目標，設定入門級與基礎級通過門檻。

設定進階高階級或流利精通級測驗之通過門檻時，程序同上述八個步驟。

完成測驗通過門檻設定後，需檢視標準設定結果是否可靠，具有效度。一般來說，標準設定結果的效度檢核可分為程序性效度、內部效度，以及外部效度三部分(Kane, 1994)，在此提供程序性效度及內部效度檢核結果。

首先，程序性效度方面，每次標準設定會議皆按照固定流程進行，且會議後的問卷調查結果顯示(見附件 1)，成員大多同意會議帶領者對於標準設定方法的操作流程說明得很清楚、每回合後團體中的討論和分享，有助於進行下一回合的判斷、對於自己所設定的切截分數有信心等等，可做為程序性效度依據。

內部效度則由每一回合估計考生能力值之標準差，及成員判斷試題難度與實徵難度參數之積差相關作為依據。標準差部分，由表 3 可知，六個等級第二與第三回合的標準差均低於第一回合。表示經過討論，判斷結果的一致性逐漸提高。從相關分析結果可發現，第三回合閱讀測驗六個等級的相關係數介於.492 至.816 之間($p<.05$)，顯示成員判斷試題難度與實徵難度間具有中度至高度的相關。上述二項分析結果均顯示標準設定結果具有內部效度。

表 3 標準設定各回合判斷結果之標準差

測驗等級	通過等級	第一回合	第二回合	第三回合
入門基礎級	入門級	0.709	0.396	0.377
	基礎級	1.282	0.471	0.511
進階高階級	進階級	0.268	0.131	0.132
	高階級	1.266	0.467	0.402
流利精通級	流利級	0.862	0.643	0.543
	精通級	1.483	0.457	0.465

華語文閱讀測驗入門基礎級、進階高階級與流利精通級三等試卷的標準設定結果，在程序性效度、內部效度二項效度證據上均獲得支持，即驗證了華語文閱讀測驗能有效將華語學習者閱讀能力區分為三等六級。

另藉由上述標準設定結果，自本年度起，提供正式考試受測者的測驗結果與等級通過判定的依據，改以量尺分數(scale score)呈現，此舉能更有效地確保測驗通過標準，不受各次組卷試題間的些微難度變化影響。本測驗之量尺分數由測驗專家基於過去參與本會所舉辦閱讀測驗之受測者能力值分布情形，並考量制定量尺分數所需考量的要素(Dorans, 2000)所訂定。三等測驗之量尺分數滿分均為 80 分，各等級通過分數如表 4 所示。

表 4 測驗通過分數

測驗等級	證書等級	通過分數
入門基礎級	入門級	42
	基礎級	60
進階高階級	進階級	48
	高階級	64
流利級 ³	流利級	52

³ 因本年度尚未辦理精通級正式考試，故僅提供流利級量尺分數門檻。

三、 測驗標準化流程

2013 年度本測驗標準化流程(如圖 1 所示)主要包含兩個部分，第一部分為正式考試製卷流程，包含：試題收集、修審、預試、分析、題庫輸入、組成正式卷、檢核正式卷與多媒體檔案等七個步驟；第二部分為考試後之成績公布，茲說明如下。

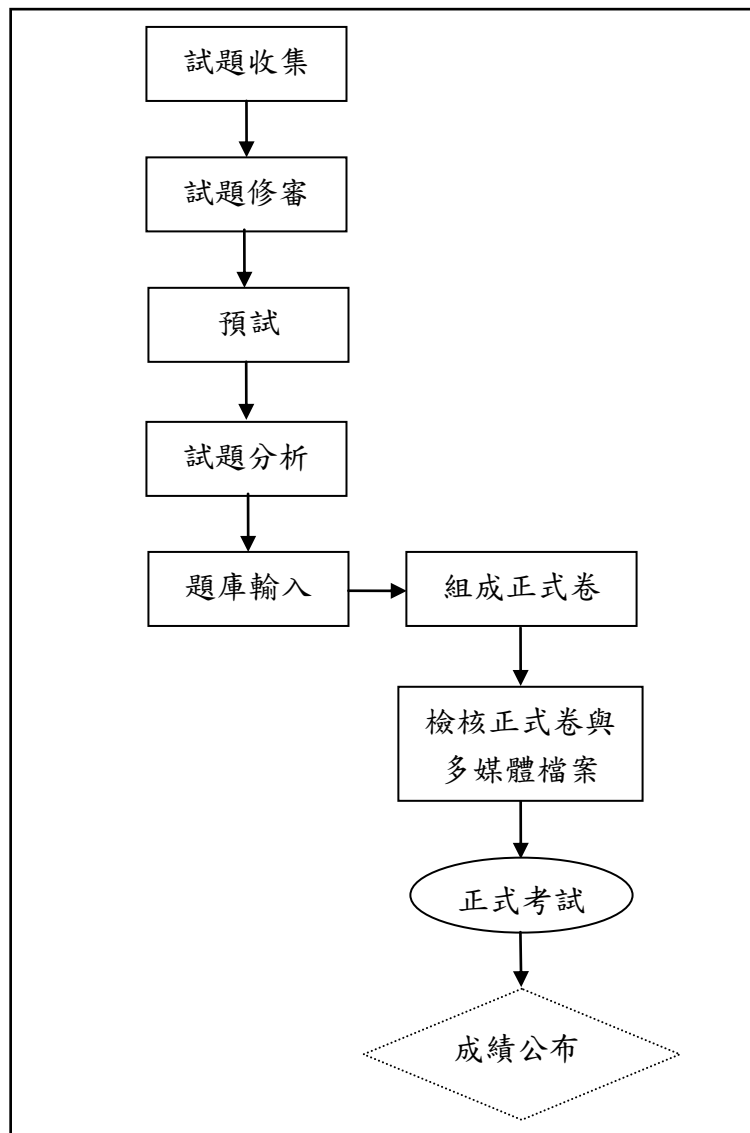


圖 1 正式考試標準化流程

(一) 試題收集

2013 年協助閱讀測驗各等級命題工作之教師共有 16 位，詳細之收題流程為：先由命題教師分期寄回所寫試題，每期再由閱讀測驗研發人員(以下簡稱研

發人員)進行審題，隨即提出反饋意見，並與命題教師進行修改討論。待確定收題後，再將需繪製圖片的試題交由專業繪圖人員進行圖片繪製，繪圖時間約為三週。本年度命題回收數量共計 836 道試題，分別為入門基礎級 327 題、進階高階級 189 題、流利精通級 320 題。

(二) 試題修審

1. 會內初審

完成試題收集後，經初步修改的試題，將交由本會非閱讀測驗之研發人員(約三人)進行第一階段會內初審，逐一完成每道試題編寫是否合宜、試卷難度分布是否合理等項目審查之後，提出審題意見，審題時間約為兩週。研發人員旋即根據會內初審意見進行試題再修改。如此，方為完成第一階段的審題工作。

2. 專家學者外審

第二階段之審題工作，乃由多名語言教學專家及資深華語教師，分別針對第一階段會內初審修改後的試題，依據試題所屬等級之雙向細目表，詳細審查試題編寫是否合宜、是否具測驗效度及難度是否符合測驗研發目標等方面，提出寶貴的審查意見。待回收專家學者審題意見後，研發人員將再依據第二階段審題意見修改試題。

3. 會內複審

針對第二階段專家學者外審修改後試題，由研發組組長進行第三階段會內複審並提供審查意見。此階段主要在逐一審查每道試題編寫是否合宜、答案分布比例是否平均等。最後，再由研發人員根據組長審題意見修改試題。

4. 製作試題相關媒體檔案

本階段將經由上述審題三步驟完成之定稿，進行試題送繪與繪圖定稿工作。

(三) 預試

經過修審後的試題皆須進入預試階段完成樣本收集程序，以獲得量化數據來客觀評估試題品質。本年度之預試樣本分別來自臺灣地區及海外地區，總施測預試題數為入門基礎級 200 題、進階高階級 200 題及流利精通級 200 題；受測人數分別為入門基礎級 1,605 位、進階高階級 2,227 位及流利精通級 1,046 位。

(四) 試題分析

將預試階段受測者之作答反應交由統計分析人員進行試題分析，並以試題反應理論(IRT)作為分析取向。由於本測驗中各試題作答反應皆為非對即錯，屬於二元計分方式(dichotomous items)，故採取 IRT 中的 Rasch 模式(如公式 1 所示)進行資料分析。由公式(1)可發現，Rasch 模式的假設為，受測者答對每一道試題機率受到受測者具備的能力及試題難度的影響，將受測者在各試題上的反應，透過公式(1)即可估計出受測者的能力以及試題的難易度，並且讓施測不同試題的受測者其能力可以互相比較。因為，在 IRT 中，受測者的能力參數與試題參數是同時納入在一個模式裡，因此，估計受測者能力時，已經考量了試題參數的影響，估計試題參數時，也考量了受測者能力的影響。

本測驗以 Winsteps 3.68.2 版分析試題品質，評估指標主要依據輸出報表中的統計指標——訊息加權適配度統計量(inlier-pattern-sensitive fit statistic)之均方差(mean-square)，及其標準化 Z 值(z-standardized)，分別簡稱為 Infit MNSQ 及 Infit ZSTD。評估標準為試題之 Infit MNSQ 數值介於 0.7 至 1.3 者，或 Infit ZSTD 介於-3.0 至 3.0 者，表示試題適配，意即試題品質與測驗研發目標一致、試題品質良好。同時，提供試題鑑別度指標⁴、難度 P 值⁵及選項分析等資料作為輔助，以更全面的觀點評估試題品質。

(五) 輸入題庫

預試後的試題在經由試題分析後，由研發人員基於各評估指標、試題相關資料及內容等進行討論，以決定試題是否予以保留。扣除共同試題後，本年度輸入閱讀測驗題庫數量總計為入門基礎級 108 題、進階高階級 207 題及流利精通級 146 題，共計 461 題。

(六) 組成正式卷

閱讀測驗正式卷由題庫之試題組成，並依照固定比例分配各難度區間之題數，自題庫中挑選難度參數符合試卷難度區間要求的試題。2012 年本會將基礎級試題難度垂直等化至進階、高階與流利級測驗難度量尺(藍珮君、陳柏熹，

⁴ 試題鑑別度指標為 D 值及點二系列相關係數。

⁵ 難度 P 值為高低分組受測者答對率之平均。

2014)，再以此難度量尺為基準向下研發入門級試題，華語文閱讀測驗試題難度量尺始為同一量尺。各等級主要難度區間組卷範圍分別為：入門基礎級介於-2.99至0.50之間；進階高階級介於-1.49至2.50之間；流利級介於-0.99至3.00之間。

正式卷初稿完成後，即交由研發人員檢核試題之主題分布是否過度集中、是否出現重複考點、是否符合雙向細目表之規劃等項目，並提出建議，再根據檢核建議更換試題，直至整份測驗卷皆符合上述要求後始得定稿。2013年各等級正式考試卷之難度分布如表5所示。

表5 各等級正式考試卷難度分布

難度	入門基礎級		進階高階級		流利級	
	I	II	I	II	I	II
<-2.99	42.0%	42.0%	-	-	-	-
-2.99~-2.50	12.0%	14.0%	-	-	-	-
-2.49~-2.00	10.0%	8.0%	-	-	-	-
-1.99~-1.50	4.0%	12.0%	2.0%	-	-	-
-1.49~-1.00	10.0%	10.0%	8.0%	20.0%	-	2.0%
-0.99~-0.50	14.0%	10.0%	38.0%	22.0%	-	-
-0.49~-0.00	4.0%	2.0%	6.0%	16.0%	-	6.0%
0.01~0.50	2.0%	2.0%	8.0%	4.0%	4.0%	2.0%
0.51~1.00	2.0%	-	12.0%	6.0%	10.0%	14.0%
1.01~1.50	-	-	10.0%	16.0%	36.0%	30.0%
1.51~2.00	-	-	6.0%	12.0%	36.0%	26.0%
2.01~2.50	-	-	10.0%	4.0%	10.0%	10.0%
2.51~3.00	-	-	-	-	4.0%	10.0%
>3.00	-	-	-	-	-	-

註：I表示第一次正式考試；II表示第二次正式考試。

(七) 檢核正式卷與多媒體檔案

首先，由研發人員登入題庫系統檢查正式卷中所有試題，主要檢核標準為：同一測驗卷中考點不重複、不互斥，題幹主題多元，以及圖片信息是否清晰等。待確認無誤後，研發人員即轉出卷包並派送至考試系統進行模擬測試，再交由本會非閱讀測驗研發人員進行相同步驟之交叉檢核，經確認無誤後，即可備試。

正式考試結束後，待檢核受測者作答資料確認無誤後，進行成績報表檢核，再次確認無誤後，始確定受測者最終考試成績，並進行成績單及證書之印製與寄發工作。

四、測驗評估

一份測驗是否能夠發揮效用，確切地測量到目標語言能力，可透過分析測驗之信度與效度來進行整體性評估。因此，本節將以 2013 年華語文閱讀測驗正式考試之信效度，來評估、總結年度之測驗效能。2013 年度本測驗正式考試分別於 5 月 4 日、11 月 2 日舉行。五月份之報考人數分別為：入門基礎級 398 名、進階高階級 896 名及流利級 347 名。十一月份報考人數則分別為：入門基礎級 255 名、進階高階級 742 名及流利級 279 名。以下分析結果，皆分別以 I 和 II 表示第一次及第二次正式考試。

(一) 信度

信度指的是測驗結果的穩定性與一致性，若一份測驗無論在任何時間、任何地點、由任何人進行施測、由任何人計分時，都能得到一致性很高的測驗結果，就表示此份測驗具有很高的信度，也就表示該測驗所獲得的測驗結果(或成績)之測量誤差很小(或精準性高)。以下將從兩大測驗取向—古典測驗理論(classical test theory；簡稱 CTT)及 IRT—分別說明本年度「華語文閱讀測驗」信度。

古典測驗理論是以測驗的信度來評估測量精準度，其假設為接受相同測驗的受測者具有相同測量精準度，因此，測量精準度是以測驗為單位來表示，即施測同一份測驗的受測者其測驗精準度皆相同。CTT 常用來評估測驗信度的指標主要分為四類，分別為：

1. 再測信度，表示在不同時間點施測所獲得的測驗成績是否具有的一致性；
2. 複本信度，表示經由施測不同題本所獲得之測驗成績是否具有穩定性；
3. 內部一致性信度，表示測驗所測量到的潛在特質是否具有的一致性；
4. 評分者信度，表示經由不同評分者所得到的評分結果是否具有的一致性。

而 IRT 的假設則為同一份測驗中的試題，對於不同能力的受測者具有不同測量精準度；當施測的試題難度越符合受測者能力時，對於獲知受測者能力越有幫助，測量精準度越高，反之，當施測試題難度與受測者能力差距越大時，對於了解受測者能力的幫助並不大，測量精準度較低。在 IRT 中，是以訊息量(information)的觀點來表示測量精準度，訊息量越高，表示施測試題對於受測者能力的測量精準度越高。

1. 內部一致性信度

由於本測驗所有受測者於同一時間接受同一測驗，因此，可藉由內部一致性信度來檢視測驗信度。當每道試題測量結果的相關性高時，則顯示測驗試題皆測量到相同的潛在特質，即內部一致性高；反之，當測量結果的相關性低，則表示測驗試題測量到其他潛在特質，即內部一致性低。

常用於表示內部一致性信度的指標有：折半信度、Cronbach's α 係數及庫李 (Kuder-Richardson) 20 號或 21 號公式。本節將分析 2013 年華語文閱讀測驗正式考試之信度，採用 Winsteps 3.68.2 版所輸出之庫李 20 號公式之係數作為內部一致性信度指標。

閱讀測驗入門基礎級、進階高階級與流利級測驗兩次正式考試之內部一致性信度如表 6 所示，由下表可知，在第一次正式考試中，三等級測驗信度係數均介於 .87 至 .91 之間；在第二次正式考試中，各等級測驗信度係數介於 .88 至 .93 之間。由此可知 2013 年華語文閱讀測驗正式考試卷信度良好。

表 6 內部一致性信度摘要表

正式考試	入門基礎級	進階高階級	流利級
I	.87	.87	.91
II	.90	.88	.93

2. 測驗訊息量

在 IRT 中，訊息量的計算公式為：

$$I_i(\theta) = \frac{P_i'(\theta)^2}{P_i(\theta)Q_i(\theta)} \quad (2)$$

其中， I_i 表示試題 i 對於能力為 θ 的受測者所提供的訊息量； $P_i'(\theta)$ 為 $P_i(\theta)$ 對 θ 的一階微分； $P_i(\theta)$ 如公式(1)所示；而 $Q_i(\theta)$ 則為 $1 - P_i(\theta)$ 。將一份測驗中所有施測試題之訊息量加總，即為測驗訊息量。測驗訊息量亦表示為測量誤差 (standard error；簡稱 SE)，如公式(3)所示：

$$SE(\theta) = \frac{1}{\sqrt{TI(\theta)}} \quad (3)$$

其中， $TI(\theta)$ 為整份測驗對於能力值為 θ 的受測者所提供的測驗訊息量。由公式(3)可知，當測驗訊息量越大，對於測量受測者能力的誤差越小，測量精準度

越高。當受測者能力分布近似於常態分布(normal distribution)時，經由 CTT 中測量標準誤與信度的關係式以及公式(3)所獲得的 SE，可換算成相對於 CTT 中的測驗信度，如公式(4)所示：

$$SEM = \sigma\sqrt{1-r} \quad (4)$$

其中，SEM 表示測量標準誤(standard error of measurement；簡稱 SEM)，即為公式(3)所獲得之 SE； σ 表示受測群體能力的標準差； r 表示測驗信度。

圖 2、圖 4 及圖 6 為入門基礎級、進階高階級和流利級測驗之測驗訊息量，而圖 3、圖 5 及圖 7 則為三等測驗之測量誤差。由圖 2 及圖 3 可知，入門基礎級測驗第一次正式考試的測驗訊息量在量尺分數為 42 分時最大，數值為 9.26，測量誤差為 0.33，而在入門級與基礎級門檻的測驗訊息量分別為 9.26 與 7.05，測量誤差分別為 0.33 與 0.38，信度係數分別為.92 與.90；第二次正式考試之測驗訊息量同樣以量尺分數 42 分為最大，數值為 9.69，測量誤差為 0.32，在入門級與基礎門檻的測驗訊息量分別為 9.69 與 7.15，測量誤差分別為 0.32 與 0.37，信度係數分別為.94 與.91。上述結果顯示，入門基礎級測驗在兩門檻所提供的測量相較於其它分數而言，具有較高或最高的訊息量與較低的測量誤差，且在兩門檻均具有良好信度。

同樣的現象在進階高階級測驗與流利級測驗(如圖 4 至圖 7)也可發現，兩等級測驗對於提供門檻之量尺分數的測驗訊息量都較高，測量誤差都較低，且信度係數介於.88 至.94 之間，具有良好信度。如進階高階級測驗之兩測驗門檻在兩次正式考試的測驗訊息量各為 9.69 與 7.65，以及 9.57 及 7.52，測量誤差均為 0.32 與 0.36，信度係數分別為.89 與.87，及.91 與.89。而流利級門檻在兩次正式考試的測驗訊息量則分別為 10.78 及 10.05，測量誤差為 0.30 及 0.32，信度係數為.93 及.94。

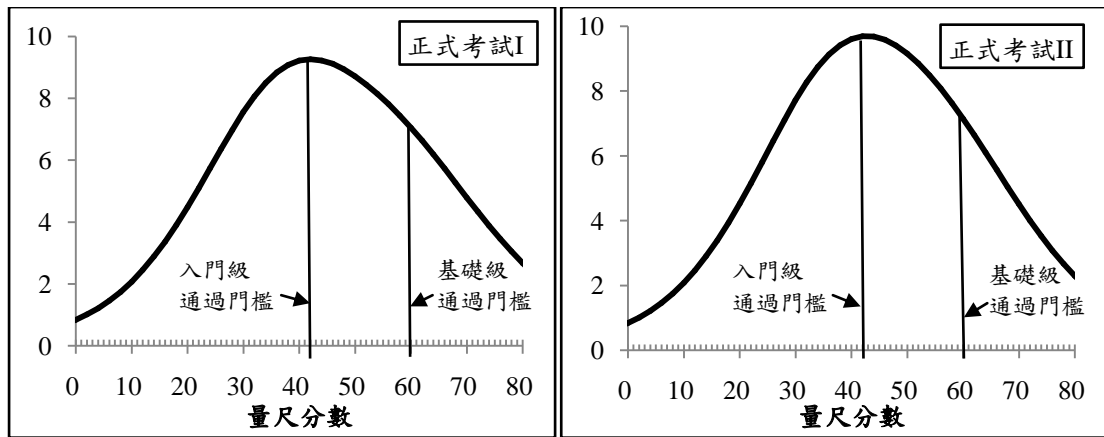


圖 2 入門基礎級測驗訊息量

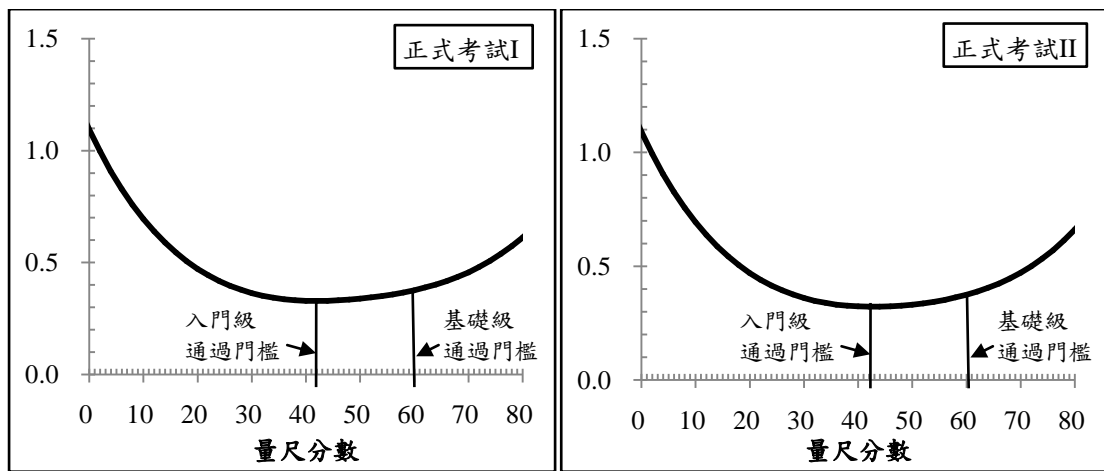


圖 3 入門基礎級測驗測量誤差

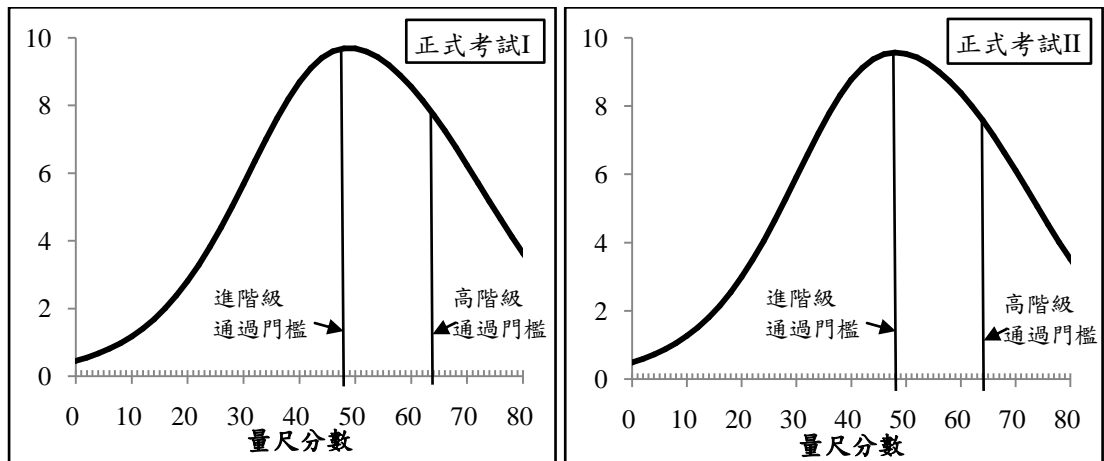


圖 4 進階高階級測驗訊息量

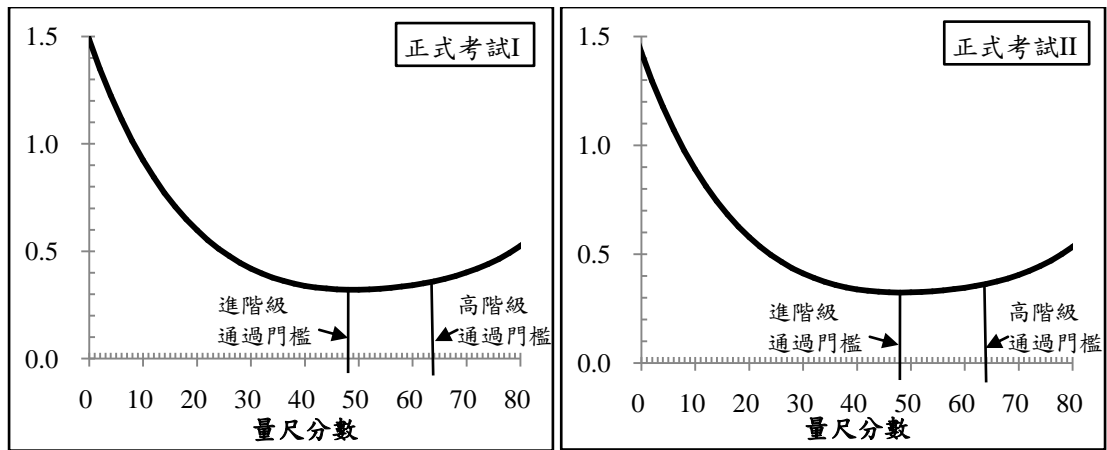


圖 5 進階高階級測驗測量誤差

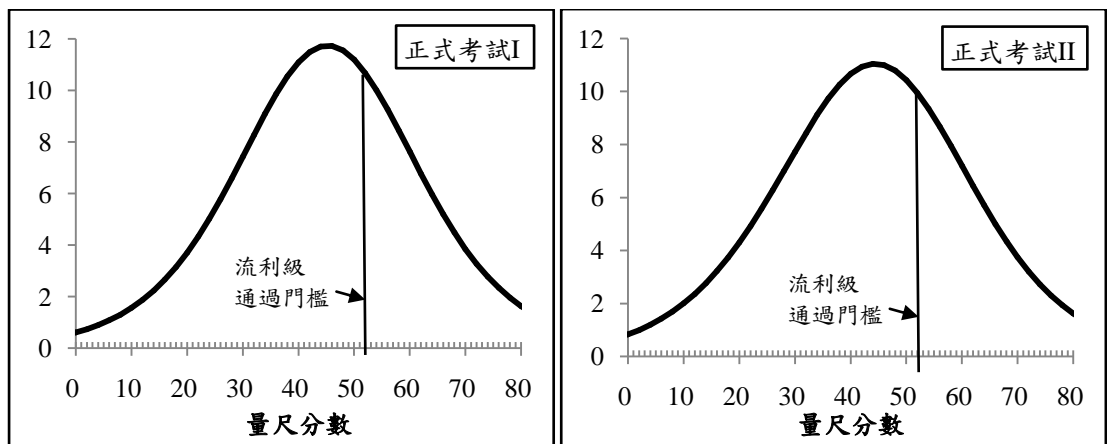


圖 6 流利級測驗訊息量

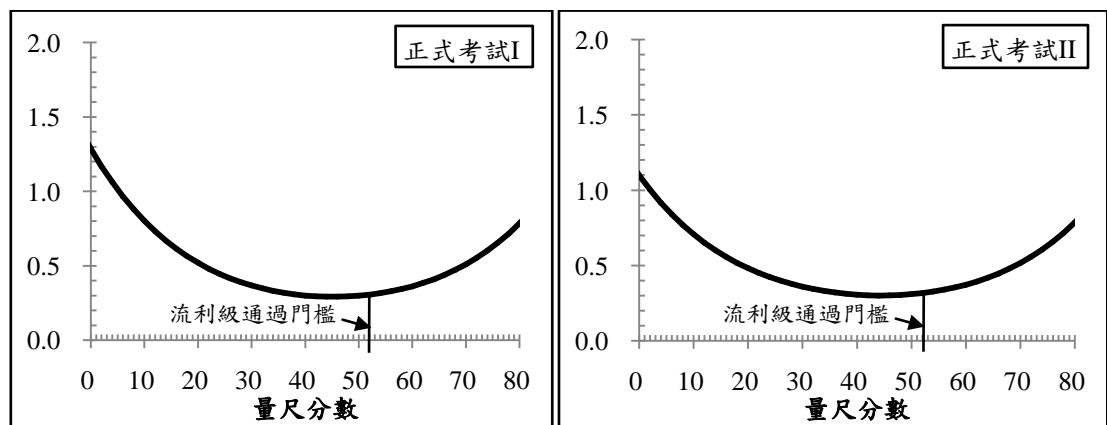


圖 7 流利級測驗測量誤差

(二) 效度

測驗效度指的是測驗是否能測量到欲測量的能力(或潛在特質)。由於欲測量的能力無法直接觀察，因此，皆須藉由受測者在試題上的作答反應或行為來間接

推估。常作為測驗效度的證據主要可分為三大類，分別為：內容效度(content validity)，指的是測驗內容的相關證據；建構效度(construct validity)，即指關於測驗架構的證據；效標效度(criterion validity)，是指測驗結果預測力的相關證據。

關於內容效度的相關證據，主要為評估和分析測驗所測量的能力及各內容題數的比重分配，與測驗所欲測量的能力定義是否相符。評估時通常採取的辦法為邀請專家評估試題是否符合上述要求。這種經由專家提供試題內容效度評估意見的方式，稱為專家效度；關於建構效度相關證據，主要在評估測驗試題所測量到的能力是否與測驗發展時所定義的架構或內容相吻合，試題分析和因素分析常被用來提供建構效度方面的證據。效標效度的相關證據則可從收集並分析受測者之相關資料與測驗結果而得。本節試從專家效度、試題分析及因素分析之驗證性因素分析(confirmatory factor analysis)三個方面，來描述 2013 年華語文閱讀測驗之內容效度及建構效度。

1. 內容效度

本會於發展閱讀測驗之初，集結了華語相關領域專家共同擬訂各等級測驗目標，並依據各等級目標訂立雙向細目表(如表 7)。命題教師通過培訓後，則嚴格依循雙向細目表及相關命題原則進行命題工作。而由研發人員回收之各等級測驗試題，皆會經過正式考試製卷流程(如圖 1)中試題修審步驟的第二階段——專家學者外審。此一階段乃由本會聘請語言教學專家及華語教師數名，審查幾條核心項目，如：確認各等級測驗試題內容是否符合該等級通過者能力描述(如表 1)、確定試題詞彙使用適切性和試題(包含題幹、選項及提問描述)表達適當性。若出現不符規定者，則於外審修審稿中進行意見標註。研發人員在回收專家學者外審意見後，針對不符合該等級審查內容之試題進行再檢視與再修改，最後將試題組合成預試卷，並逐步完成預試、預試收樣、量化資料蒐集和試題品質分析等工作。

表 7 雙向細目表

分項能力	題型 測驗等級	單句理解	看圖釋義	完成段落	選詞填空		材料閱讀	短文閱讀	
		入門基礎	入門基礎	入門基礎	入門基礎	進階高階	流利	進階高階	進階高階
整體性的閱讀理解				V		V	V		
語法能力		V			V				
閱讀書信							V		
導向閱讀			V				V		
為資訊論證而閱讀			V					V	V
閱讀說明指導							V		

2. 建構效度

(1) 試題分析

由於本測驗組卷方式是依據 IRT 而來，而 IRT 的一項重要假設為：單向度假設。所謂單向度假設，指的是測驗中所有試題皆測量相同潛在特質，也就是說，受測者回答試題皆仰賴單一特質。在此一前提下，我們即可藉由試題分析來評估測驗效度，若每道試題分析結果皆符合單向度假設，即表示所有試題皆測量到相同潛在特質，以此做為測驗建構效度之證據。為此，本節將採用 Winsteps 3.68.2 版中的 Rasch 模式(如公式 1 所示)分析資料，結果詳述如下。

整體而言，在試題難度分布上，各等級兩次正式考試的試題難度分布相近，例如，難度平均值之絕對值差異皆小於 0.15 logit、標準差的絕對值差異小於 0.3 (如表 8)。能得出此一結果主要是由於各等級測驗在組合正式卷時，均依據各等級測驗主要難度區間組卷範圍而來(如表 5)。

在考生能力分布方面(參見表 9)，總體而言，各等級測驗在兩次正式考試中，考生平均能力值之絕對值差異均小於 0.2 logit；各等級考生平均能力估計標準差的絕對值差異均小於 0.2。透過上述結果可知，各等級測驗在兩次正式考試中，考生平均能力分布大致相近。

在入門基礎級兩次正式考試中，受測者能力估計值分別為-0.17 與-0.05，50 道試題的平均難度估計值分別為-2.38 以及-2.48。以上結果顯示，兩次正式考試之受測者平均能力與試題平均難度絕對值差異分別為 2.2 及 2.4 logits，相對於受測者而言，試題較為容易；同樣的現象可由圖 8 受測者與試題分布圖(person-item map)觀察到，我們將受測者能力估計值與試題難度估計值放在一起(試題置於圖

的右方、受測者能力置於圖的左方)，自上而下，將受測者能力由高到低、試題難度由難到易進行排列，發現試題難度涵蓋範圍與受測者能力分布範圍差距較大，即測驗的試題對受測者來說較為簡單。造成受測者平均能力與試題平均難度的絕對值差異較大的原因可能為：在臺灣地區參加正式考試的受測者，每天長時間沉浸於目標語環境中生活與學習，學習成效較高，因此可以在較短時間內具備基礎的閱讀能力表現。兩次正式考試試題平均難度 95% 信賴區間所涵蓋的受測者能力範圍分別為 58% 與 40%。

在進階高階級與流利級的兩次正式考試中，也得到與基礎級測驗類似的結果，試題相對於受測者能力而言偏易。不過，受測者平均能力與試題平均難度絕對值差異較基礎級小，乃介於 0.5 至 0.7 logit 之間。我們將受測者能力估計值與試題難度估計值(參見圖 9、圖 10)兩相對比時還可發現，大部分試題難度與受測者能力分布範圍相當，即測驗試題難度範圍可涵蓋大多數受測者能力分布範圍。透過上述結果可知，隨著測驗等級越高，考生平均能力估計值與試題平均難度估計值差異越小。進階高階級兩次正式考試試題平均難度 95% 信賴區間所涵蓋的受測者能力範圍分別為 96% 與 92%；流利級則分別為 68% 與 75%。

表 8 試題難度估計分布

測驗等級	入門基礎級		進階高階級		流利級	
	I	II	I	II	I	II
正式考試						
平均值	-2.38	-2.48	0.16	0.09	1.52	1.39
標準差	1.28	1.13	1.12	1.14	0.53	0.80
最大值	0.98	0.04	2.36	2.17	2.80	2.84
最小值	-3.83	-3.90	-1.67	-1.49	0.17	-1.01

表 9 受測者能力估計分布

測驗等級	入門基礎級		進階高階級		流利級	
	I	II	I	II	I	II
正式考試						
平均值	-0.17	-0.05	0.61	0.81	2.09	2.04
標準差	1.24	1.38	0.99	1.11	1.25	1.39
最大值	3.63	3.28	4.67	5.84	6.77	6.78
最小值	-3.22	-3.36	-2.24	-2.32	-0.39	-1.06

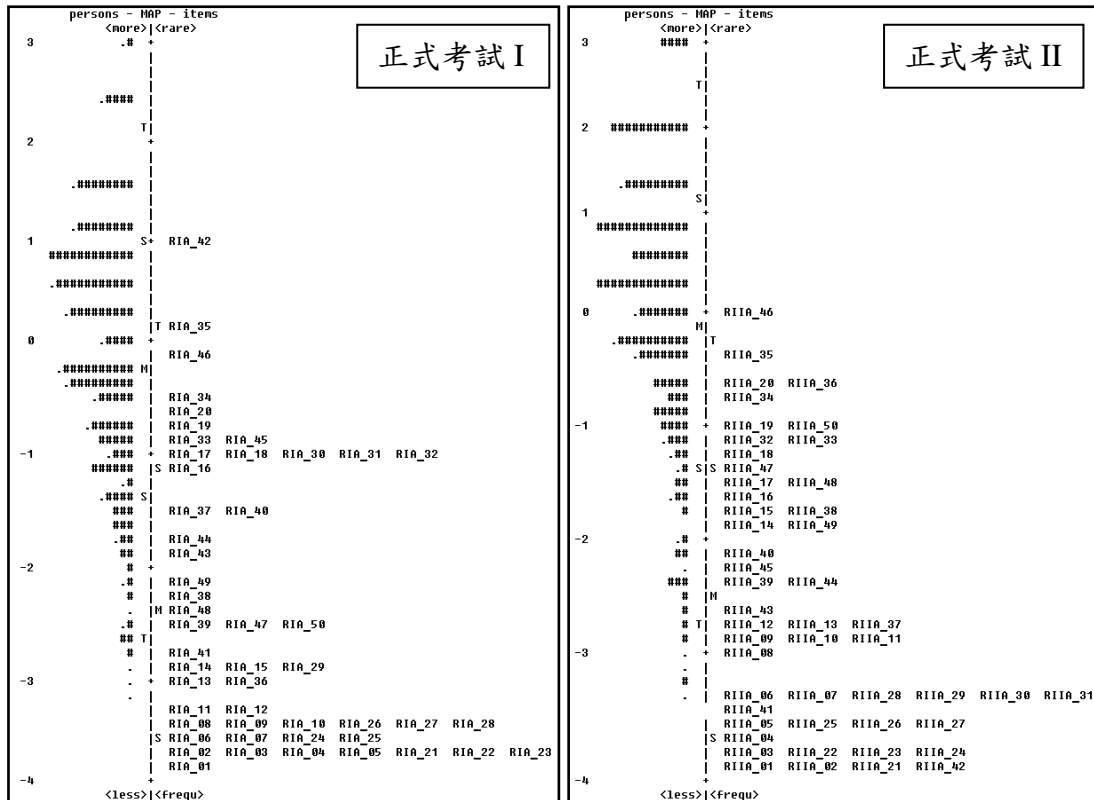


圖 8 入門基礎級測驗受測者與試題分布圖

註：“#”在正式考試 I 與 II 分別表示三位及二位受測者；“.”表示一位受測者；RIA_01~RIA_50 和 RIIA_01~RIIA_50 表示試題編號。

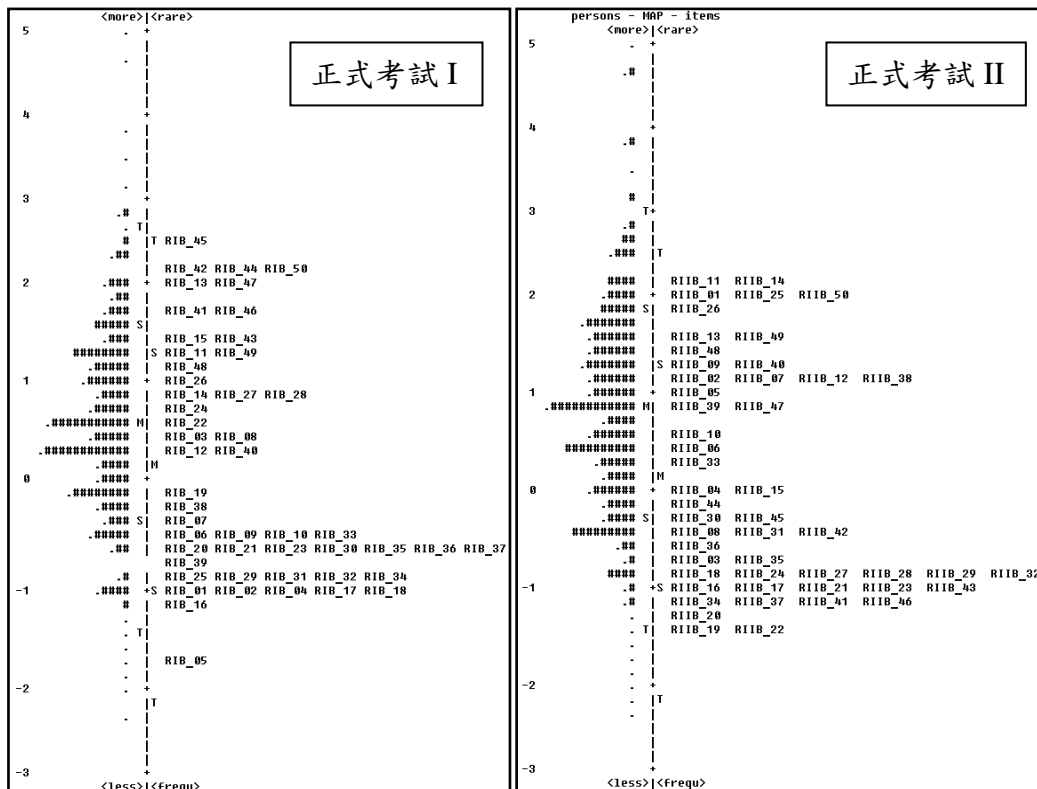


圖 9 進階高階級測驗受測者與試題分布圖

註：“#”在正式考試 I 與 II 分別表示七位及五位受測者；“.”表示一位受測者；RIB_01~RIB_50 和 RIIB_01~RIIB_50 表示試題編號。

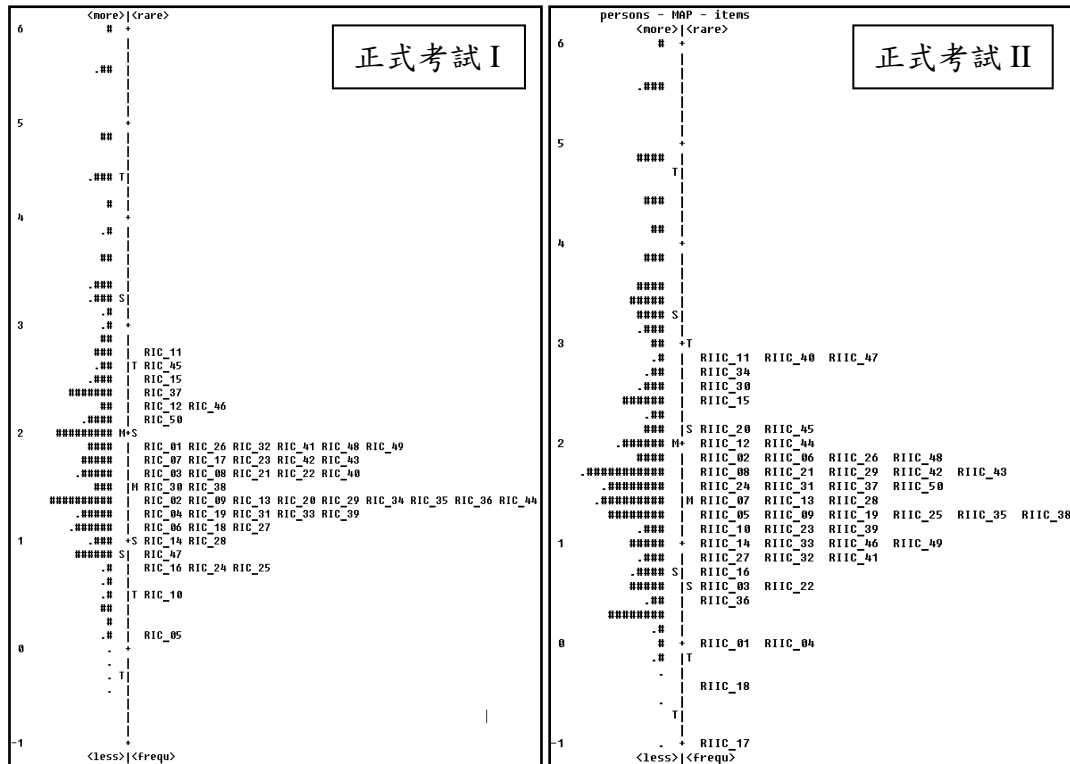


圖 10 流利級測驗受測者與試題難度圖

註：“#”在正式考試 I 與 II 分別表示三位及二位受測者；“.”表示一位受測者；RIC_01~RIC_50 和 RIIC_01~RIIC_50 表示試題編號。

採用 Infit MNSQ 介於 0.6 到 1.4 及 ZSTD 介於-3.0 到 3.0 的標準，來評估試題是否與單向度 IRT 模式適配。各等級在兩次正式考試中不符合單向度 IRT 模式的試題數量如表 10 所示。入門基礎級兩次測驗各 50 道試題中，分別有九道及四道試題；進階高階級兩次測驗各 50 道試題中，分別有三道及四道試題；流利級兩次測驗各 50 道試題中，第二次有三道試題，第一次沒有試題不適配。各等級測驗在兩次正式測驗的試題適配率介於 82% 至 100%，顯示適配比例大致良好。

一般來說，造成試題與模式不適配的原因有以下幾種，一為當受測者答題所需要的能力超過單一向度時，則可能造成試題不適配。因在 Rasch 模式的假設下，受測者答對試題機率僅受到單向度能力(此為閱讀能力)所影響。二是當試題具有其他會影響到答題機率的特性，例如，試題的猜測度或鑑別度，而未能被 Rasch 模式納入解釋時，也會發生試題不適配情形。三是當受測者因為粗心或其他原因而使其作答未能符合自身能力表現時，亦會造成試題估計不符合模式。例如，在模式假定受測者能力越高答對機率越高的前提下，高能力之受測者卻答錯較簡單的題目。

由於本測驗正式卷的試題皆經過預試，且為獲得穩定難度參數估計值後所組

成，故測驗觀察資料與測量模式不適配的原因應為第三種情形：在分析時，試題難度被設定為已知，再進行模式適配度考驗，而受測者因粗心或其他原因使得作答未符合其能力表現，造成試題估計不符合模式。

表 10 試題適配分布

測驗等級	入門基礎級		進階高階級		流利級	
	I	II	I	II	I	II
正式考試						
總試題數	50	50	50	50	50	50
適配題數	41	43	47	46	50	47
不適配題數	9	7	3	4	0	3
試題適配率(%)	82	86	94	92	100	94

進一步針對各等級測驗不適配試題進行選項分析。表 11 至 13 列出了各等級測驗不適配試題之難度參數估計值及其估計誤差、各選項受測者選答人數、選答各選項的受測者平均能力估計值及其估計誤差、點二系列相關係數等統計數值。

由表 11 可知，在入門基礎級兩次正式考試共 16 道不適配試題中，選答正確選項的人數都是最多的，該選項平均能力估計值也最高，這表示選答該選項的受測者平均能力高於選答錯誤選項者，符合模式的預期——能力越高的受測者答對試題的機率越高。再看試題中正確選項的點二系列值，除了 RIA_04、RIA_05、RIA_21、RIA_41、RIIA_13、RIIA_21、RIIA_39 低於.20 外，其餘試題相關係數介於.21 至.39 之間，這表示測驗具有幅合效度(convergent validity)，即能力越高者答對試題的機率越高。此外，從錯誤選項值較低且幾乎為負值來看，則可知測驗具有區辨效度(discriminate validity)，即可有效區別受測者能力之高低。

正確選項點二系列相關較低的七道試題，由圖 11 至 17 試題特徵曲線(item characteristic curve)可以觀察到，均有部分能力較高的受測者答錯，或部分能力較低的受測者答對。以 RIA_41 為例，此題能力值與試題難度相差在 2 至 4 之間的受測者，依照模式預期，應有 88% 至 98% 的答對率，但實際作答資料顯示，此一區間受測者答對率較低，約介於 75% 至 80% 之間，且低於 95% 信賴區間之外，表示有部分受測者答錯此題，與模式預期不符，並使得正確選項點二系列相關係數較低。

表 11 入門基礎級測驗不適配試題選項描述性摘要表

正式 考試	試題編號	難度	估計 誤差	選 項	計 分	人數	平均 能力	平均能力 估計誤差	點二系 列相關
I	RIA_02	-3.784	0.244	B	0	29	-1.51	0.17	-.31
				C	0	9	-1.50	0.24	-.16
				A	1	360	-0.03	0.06	.35
	RIA_04	-3.732	0.239	A	0	19	-1.05	0.21	-.13
				C	0	19	-0.84	0.26	-.09
				B	1	360	-0.09	0.07	.16
	RIA_05	-3.713	0.238	A	0	3	-2.57	0.33	-.21
				C	0	35	-0.45	0.14	.01
				B	1	360	-0.12	0.07	.05
	RIA_09	-3.441	0.215	B	0	4	-2.57	0.12	-.24
				A	0	9	-1.90	0.29	-.24
				C	1	385	-0.10	0.06	.33
	RIA_10	-3.440	0.215	B	0	8	-1.88	0.24	-.22
				A	0	44	-0.80	0.19	-.15
				C	1	346	-0.05	0.06	.23
	RIA_11	-3.409	0.212	A	0	2	-2.84	0.16	-.20
				B	0	54	-1.05	0.17	-.28
				C	1	342	-0.01	0.06	.31
RIA_21	-3.787	0.245	A	0	14	-0.71	0.36	-.07	
			C	0	18	-0.62	0.27	-.04	
			B	1	366	-0.12	0.06	.08	
RIA_22	-3.762	0.242	A	0	2	-2.84	0.38	-.20	
			B	0	5	-2.28	0.42	-.23	
			C	1	391	-0.13	0.06	.30	
RIA_41	-2.806	0.173	B	0	23	-0.92	0.21	-.12	
			C	0	53	-0.32	0.15	.02	
			A	1	322	-0.09	0.07	.05	
II	RIIA_10	-2.820	0.223	C	0	2	-2.63	0.31	-.21
				B	0	10	-2.07	0.20	-.33
				A	1	243	0.06	0.09	.39
	RIIA_13	-2.752	0.219	A	0	21	-0.70	0.29	-.11
				B	0	18	-0.49	0.17	.00
				C	1	216	0.05	0.10	.08
	RIIA_15	-1.789	0.174	C	0	10	-1.43	0.41	-.21
				A	0	69	-0.55	0.14	-.12
				B	1	176	0.23	0.10	.21
	RIIA_21	-3.820	0.309	C	0	22	-0.71	0.25	-.09
				B	1	233	0.02	0.09	.09
				B	0	8	-0.88	0.51	-.10
	RIIA_39	-2.351	0.197	A	0	48	-0.56	0.17	-.10
				C	1	199	0.11	0.10	.14
				C	0	49	-1.00	0.18	-.32
	RIIA_40	-2.133	0.187	B	0	36	-0.14	0.16	.07
				A	1	170	0.25	0.11	.21
				B	0	10	-1.98	0.29	-.37
RIIA_47	-1.321	0.161	E	0	16	-1.44	0.20	-.29	
			F	0	20	-0.83	0.19	-.13	
			D	0	16	-0.44	0.14	.00	
			C	0	44	-0.18	0.13	.05	
			A	1	143	0.54	0.11	.32	
			M	0	6	-2.36	0.34	-.33	

註： M 表示沒有填答。

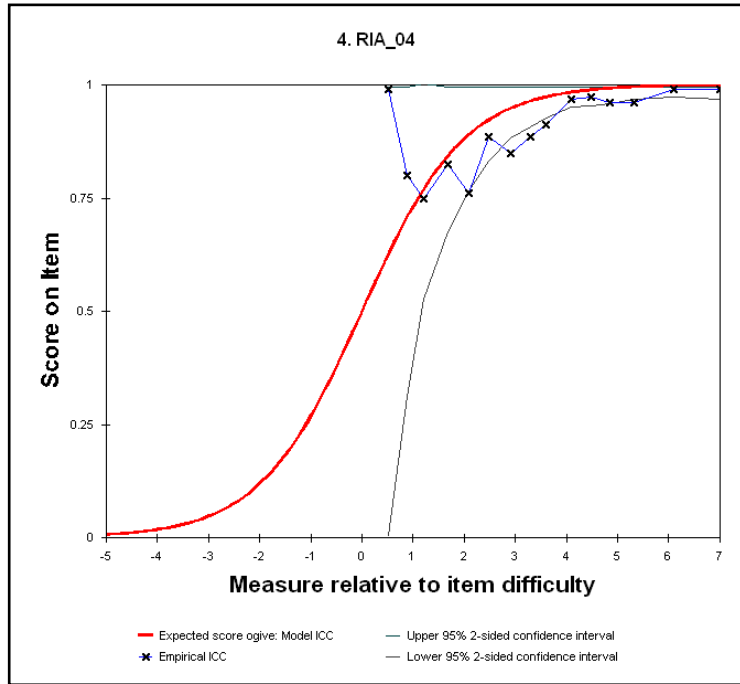


圖 11 RIA_04 試題特徵曲線

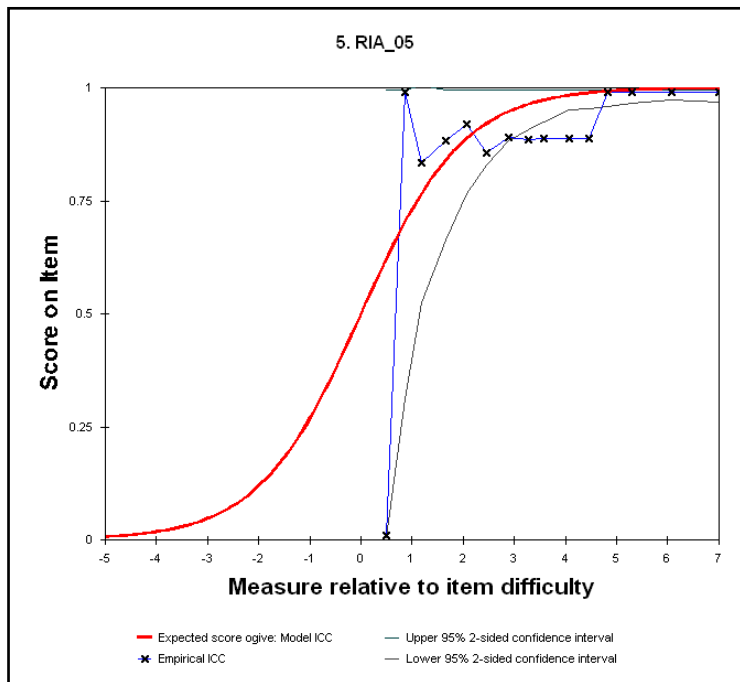


圖 12 RIA_05 試題特徵曲線

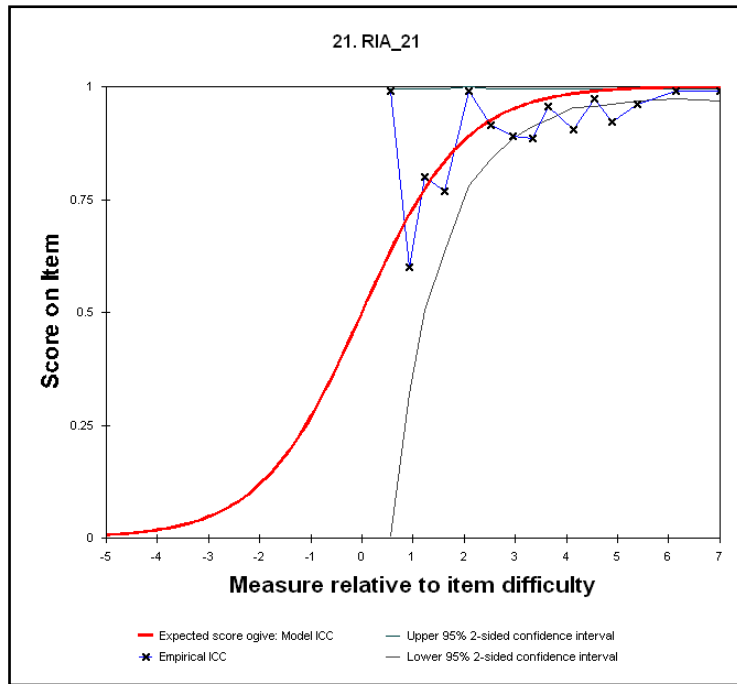


圖 13 RIA_21 試題特徵曲線

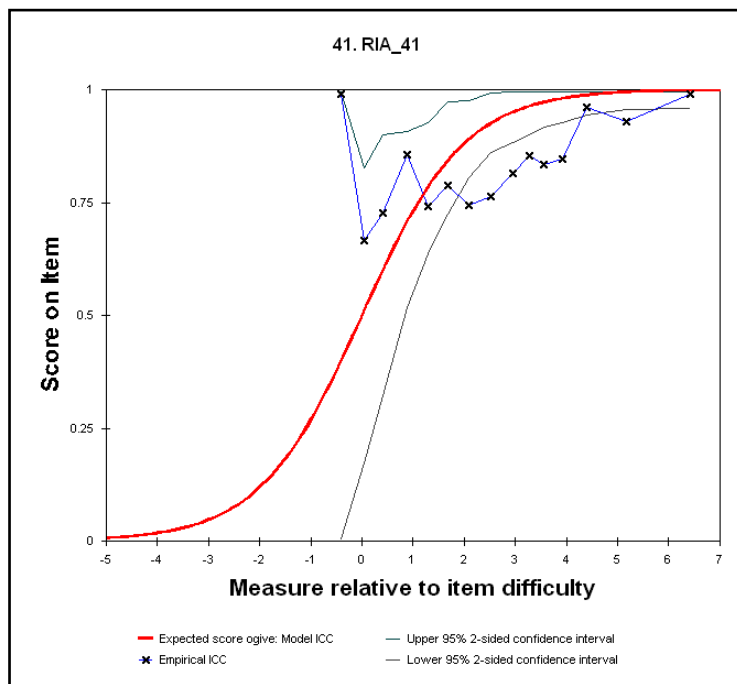


圖 14 RIA_41 試題特徵曲線

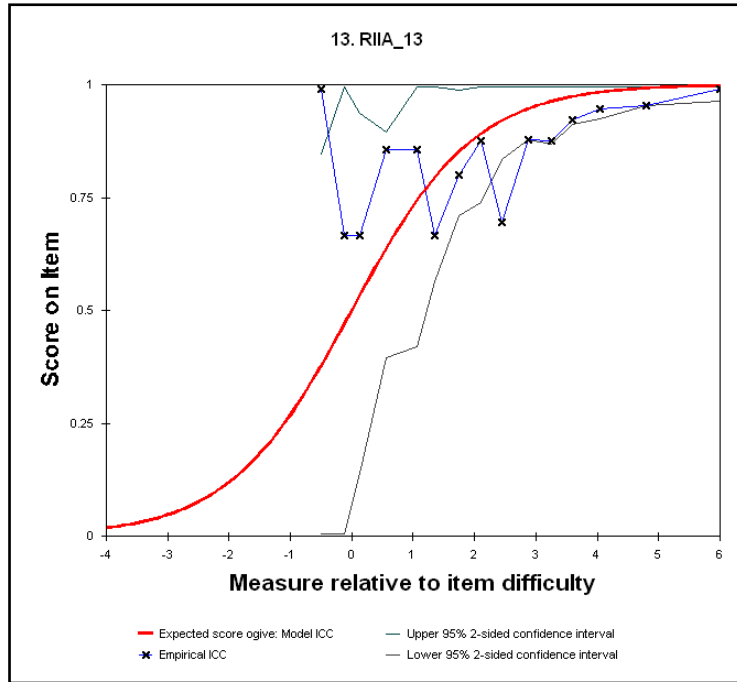


圖 15 RIIA_13 試題特徵曲線

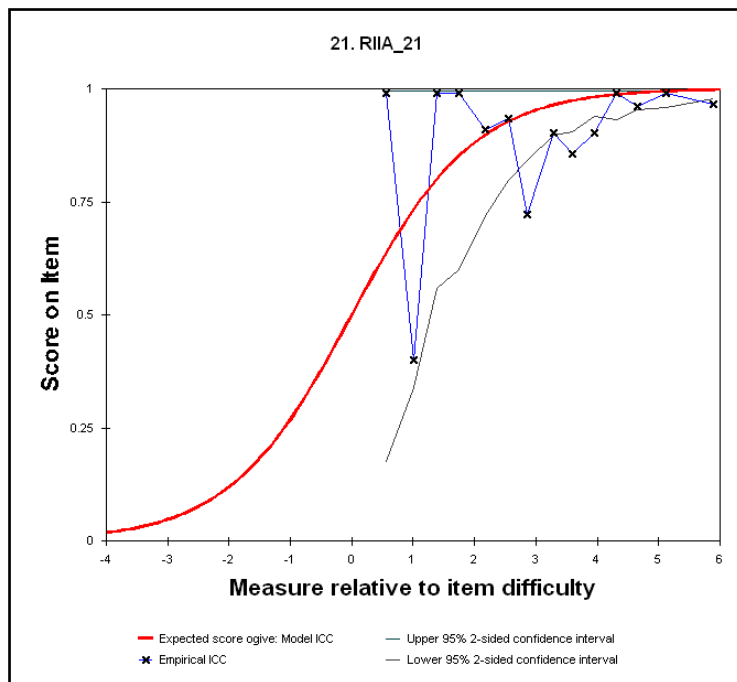


圖 16 RIIA_21 試題特徵曲線

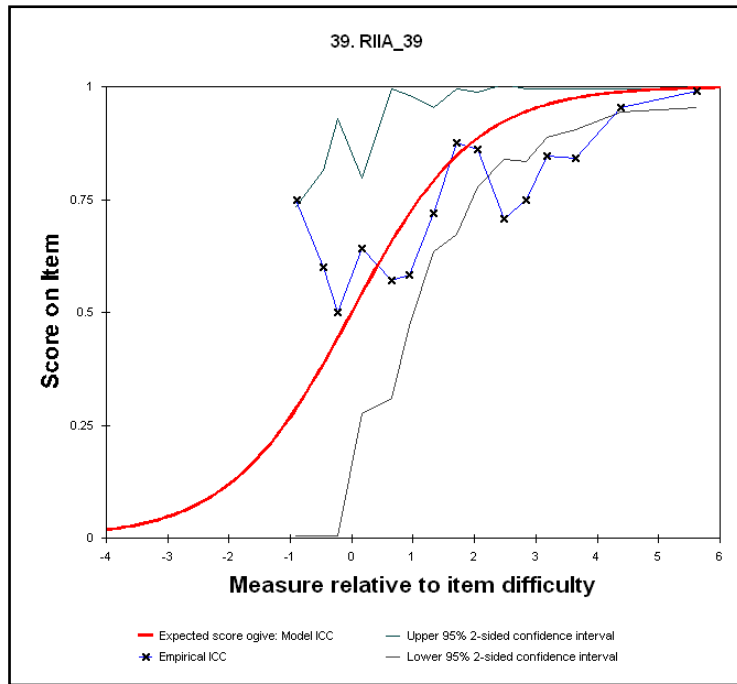


圖 17 RIIA_39 試題特徵曲線

進階高階級測驗選項分析結果如表 12 所示，與入門基礎級相仿，不適配試題選答正確選項的人數都是最多的，該選項平均能力估計值也最高，符合模式預期。在點二系列相關方面，第一次及第二次正式考試中，分別有兩道及一道試題的點二系列相關低於.20，其餘試題相關係數介於.22 至.51 之間，這表示測驗具有幅合效度，即能力越高者答對試題的機率越高。此外，從錯誤選項值較低且幾乎為負值來看，則可知測驗具有區辨效度，即可有效區別受測者能力之高低。

此三道試題特徵曲線如圖 18 至圖 20 所示，承前所述，究其可能原因為：有部分能力較高的受測者答錯，或部分能力較低的受測者答對，因此造成點二系列相關偏低。舉例來說，試題編號 RIIB_14 難度參數為 2.170，據模式預期，能力值與難度參數相減介於-3.5 至-1.5 之間的受測者答對此道試題的機率均低於 25%，然而，實際作答資料，此一區間受測者答對率均高於 25%，且超過 95% 信賴區間，因而造成該試題與模式不適配，正確選項相關係數較低。

表 12 進階高階級測驗不適配試題選項描述性摘要表

正式 考試	試題編號	難度	估計 誤差	選項	計分	人數	平均 能力	平均能力 估計誤差	點二系 列相關
I	RIB_29	-0.917	0.086	A	0	50	-0.27	0.11	-.21
				D	0	50	0.11	0.10	-.11
				B	0	224	0.25	0.06	-.17
				C	1	571	0.89	0.04	.31
				M	0	1	-1.87	--	-.08
	RIB_46	1.744	0.082	B	0	170	0.41	0.07	-.09
				D	0	210	0.44	0.06	-.07
				A	0	96	0.45	0.08	-.04
				C	1	370	0.92	0.06	.16
				M	0	50	0.17	0.14	-.09
	RIB_50	2.081	0.087	D	0	171	0.42	0.06	-.09
				B	0	208	0.44	0.06	-.08
				A	0	129	0.59	0.07	.01
				C	1	294	0.94	0.07	.14
				M	0	94	0.40	0.10	-.05
II	RIIB_14	2.170	0.094	A	0	56	0.30	0.10	-.13
				B	0	192	0.40	0.07	-.20
				C	0	183	1.02	0.07	.15
				D	1	297	1.09	0.07	.12
				M	0	14	-0.05	0.27	-.11
	RIIB_32	-0.810	0.098	A	0	12	-0.78	0.24	-.19
				C	0	41	-0.69	0.09	-.35
				D	0	35	-0.59	0.10	-.30
				B	1	652	1.02	0.04	.51
				M	0	2	-0.82	0.56	-.08
	RIIB_37	-1.103	0.105	D	0	54	-0.40	0.09	-.32
				A	0	47	-0.16	0.12	-.23
				B	0	117	0.24	0.08	-.20
				C	1	521	1.16	0.05	.47
				M	0	3	-0.20	0.89	-.06
RIIB_45	-0.355	0.090	A	0	34	0.25	0.17	-.11	
			D	0	81	0.36	0.10	-.13	
			C	0	166	0.54	0.07	-.10	
			B	1	448	1.08	0.05	.22	
			M	0	13	-0.43	0.17	-.16	

註：M 表示沒有填答；"--"表示該選項只有一名受測者選答，無法計算平均能力估計標準差。

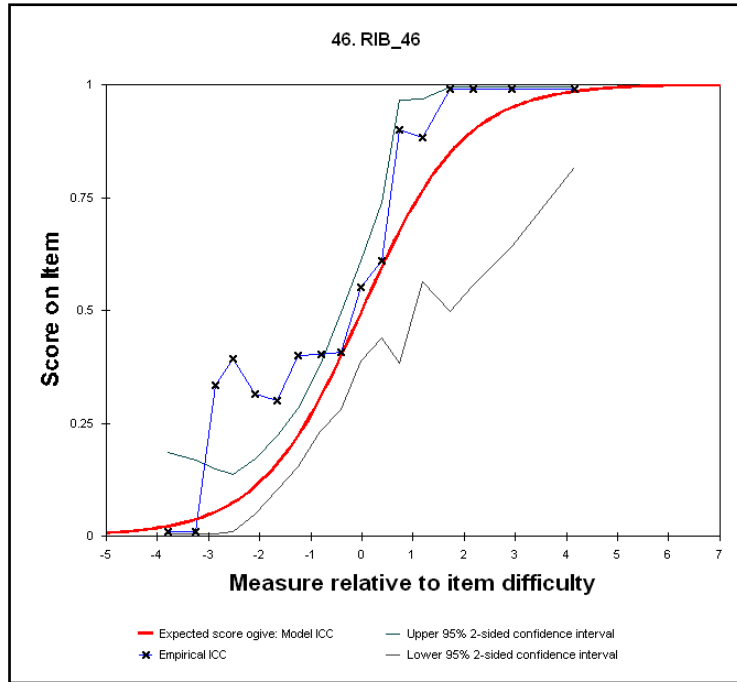


圖 18 RIB_46 試題特徵曲線

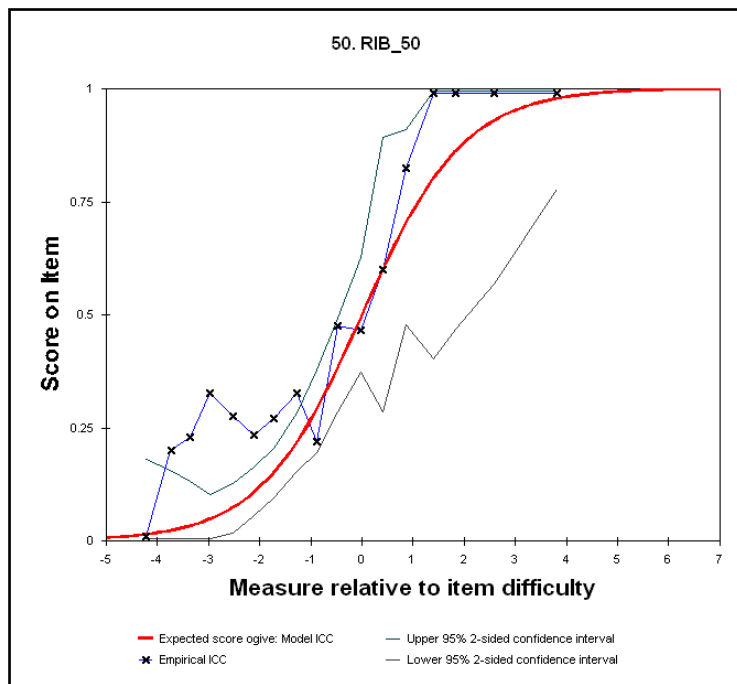


圖 19 RIB_50 試題特徵曲線

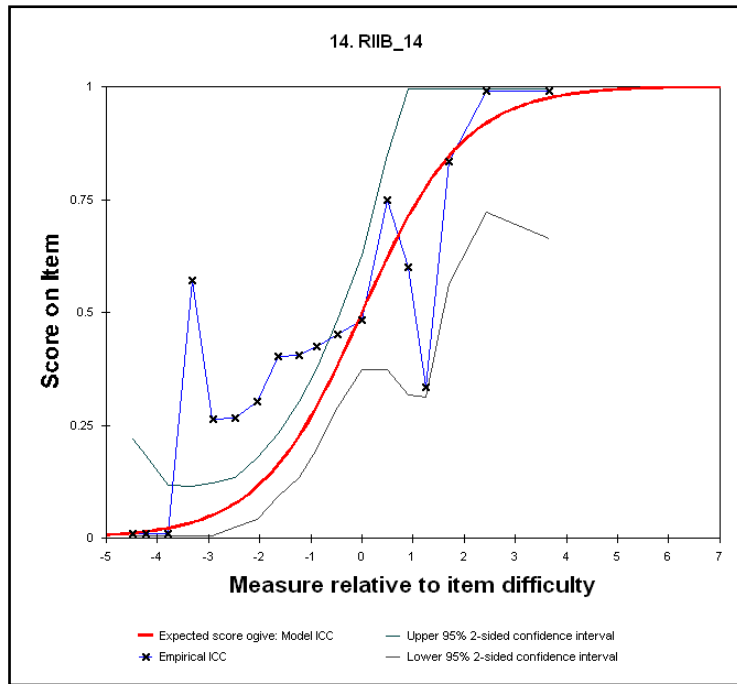


圖 20 RIIB_14 試題特徵曲線

表 13 為流利級測驗選項分析結果，三道不適配試題，選答正確選項人數都是最多的，該選項平均能力估計值也是最高的。此外，正確選項的點二系列值最高且皆為正值，此即表示測驗具有輻合效度；而錯誤選項值較低且幾乎為負值，則表示測驗具有區辨效度，即可有效區分受測者的能力高低。

表 13 流利級測驗不適配試題選項描述性摘要表

正式 考試	試題編號	難度	估計 誤差	選項	計分	人數	平均 能力	平均能力 估計誤差	點二系 列相關
II	RIIC_04	-0.001	0.170	D	0	19	0.97	0.19	-.22
				A	0	26	1.54	0.22	-.10
				C	0	22	1.59	0.23	-.05
				B	1	211	2.25	0.10	.23
				M	0	1	-0.64	--	-.12
	RIIC_11	2.796	0.148	B	0	31	1.35	0.16	-.16
				D	0	49	1.36	0.16	-.24
				A	0	91	2.09	0.14	.06
				C	1	107	2.53	0.14	.24
				M	0	1	-0.64	--	-.12
	RIIC_47	2.820	0.148	C	0	53	1.36	0.11	-.24
				A	0	39	1.49	0.18	-.15
				D	0	67	2.11	0.15	.06
				B	1	113	2.57	0.15	.24
				M	0	7	0.95	0.30	-.13

註：M 表示沒有填答；"--"表示該選項只有一名受測者選答，無法計算平均能力估計標準差。

(2) 驗證性因素分析

除了透過試題分析來評估閱讀測驗是否具有建構效度之外，本測驗亦從結構方程模型(structural equation modeling)之驗證性因素分析來評估測驗的建構效度。由於閱讀測驗題型為單選題，答題反應皆為非對即錯的二分名義變項，故本研究使用可分析二元變項和類別變項的 Mplus 分析軟體，來檢視閱讀測驗入門基礎級、進階高階級與流利級雙向細目中欲測得之閱讀分項能力，是否可經由考生於試題的答題反應測知。在每一等級之因素分析模型中，測量變數為每道試題，而潛在變數為閱讀分項能力，例如，入門基礎級測驗中，測量變數為 50 道試題，潛在變數為「整體性的閱讀理解」、「為資訊論證而閱讀」與「導向閱讀」。

此節使用前述之 Mplus 7.0 版進行資料分析，估計方法採用「平均數與變異數修正後的加權最小平方值法」(weighted least squares means and variance adjusted, WLSMV)，各等級測驗之驗證性因素分析結果則分別透過基本適配度及整體適配度指標進行模式評估。

依據 Bagozzi 和 Yi(1998)以及 Hu 和 Bentler(1998)的研究，訂定出基本適配度的評估標準如下：(1)因素負荷量宜介於.5 至.95 之間；(2)相關係數不可大於 1.0；(3)不能有過大的標準誤。至於整體適配指標部分，則採用卡方自由度比(χ^2/df)來評估整個模式與觀察資料的適配程度；以平方概似平方誤根係數指標(root mean square error of approximation；簡稱 RMSEA)來評估整體模式的絕對適配度；以非規範適配指標(non-normed fit index；簡稱 NNFI，亦稱為 TLI)與比較適配指標(comparative-fit index；簡稱 CFI)二項指標來評估整體模式增值適配度。判斷標準分別為： $\chi^2/df < 3$ 、RMSEA $< .08$ 、CFI 和 NNFI $> .90$ 。

基本適配指標部分，閱讀測驗入門基礎級、進階高階級與流利級 5 月和 11 月正式考試驗證性因素分析結果如下圖 21 至圖 23 所示，由於篇幅有限，各等級各題因素負荷量等參數請見附件 2 至 4。入門基礎級測驗三因素模型驗證性因素分析結果顯示，5 月正式考試因素負荷量介於.08 至.84 之間，平均值為.56；因素負荷量標準誤為 0.04 至 0.12；除 RIA_5、RIA_21、RIA_41 外，因素負荷量統計考驗均達顯著水準($p < .05$)；潛在變項間的相關係數為.89、.88 及.95。11 月正式考試因素負荷量介於-.07 至.96 之間，平均值為.63；因素負荷量標準誤為 0.04 至 0.12；除 RIIA_13、RIIA_21、RIIA_42 外，因素負荷量統計考驗均達顯著水準($p < .05$)；潛在變項間的相關係數為.89、.74 及.81。

進階高階級測驗四因素模型驗證性因素分析結果顯示，5月正式考試因素負荷量介於.12至.70之間，平均值為.46；因素負荷量標準誤為0.03至0.06；因素負荷量統計考驗均達到顯著水準($p < .05$)；四個潛在變項間的相關係數介於.85至.96之間。11月正式考試因素負荷量介於.15至.85之間，平均值為.52；因素負荷量標準誤為0.03至0.06；因素負荷量統計考驗均達顯著水準($p < .05$)；四個潛在變項間的相關係數介於.75至.95之間。

流利級測驗二因素模型驗證性因素分析結果顯示，5月正式考試因素負荷量介於.28至.79之間，平均值為.55；因素負荷量標準誤為0.04至0.08；因素負荷量統計考驗均達顯著水準($p < .05$)；潛在變項間的相關係數為.93。11月正式考試因素負荷量介於.34至.77之間，平均值為.60；因素負荷量標準誤為0.04至0.08；因素負荷量統計考驗均達顯著水準($p < .05$)；潛在變項間的相關係數為.87。

根據以上指標分析，可知閱讀測驗各等級兩次正式考試之驗證性因素模型基本適配度檢驗大致良好，試題因素負荷量介於.50至.95的比例在46%至84%之間，表示大多數試題均符合標準。若以一般測驗實務上可接受之因素負荷量.30為基準，入門基礎級兩次正式考試因素負荷量低於.30的試題比例皆為10%，進階高階級和流利級依序為14%、12%、2%及0%，顯示僅少數的試題因素負荷量較低，而因素負荷量較低之試題，於未來組正式卷時將特別留意。

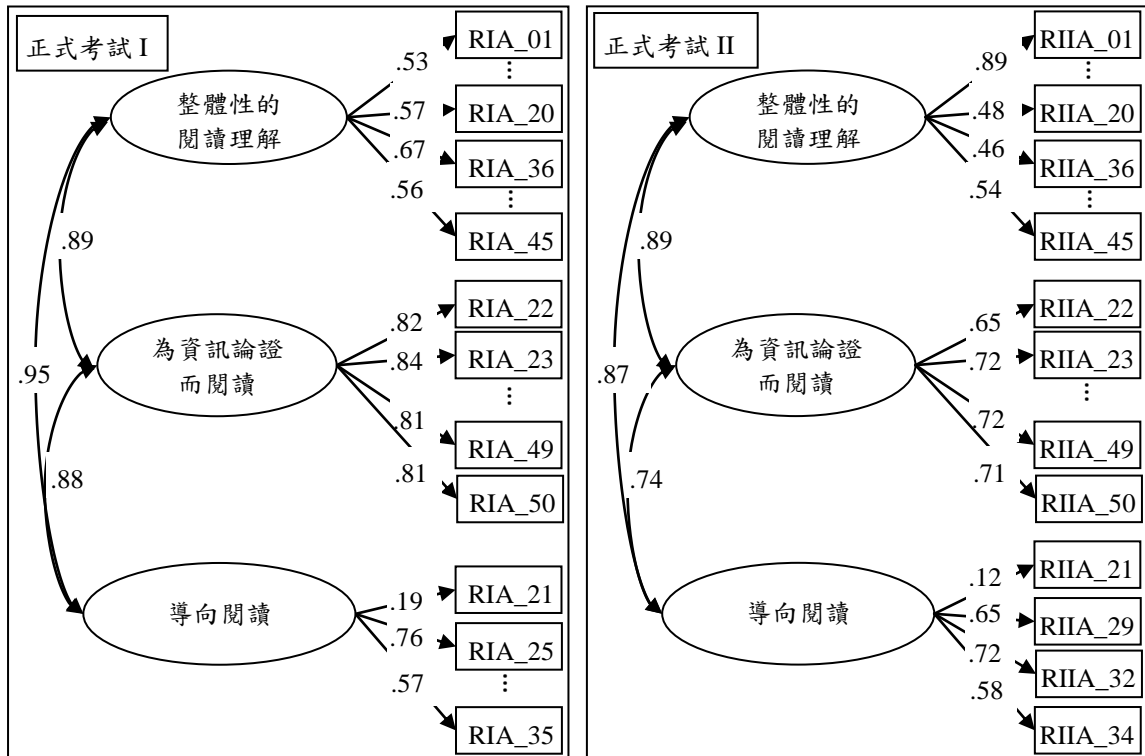


圖 21 入門基礎級閱讀測驗三因素驗證性因素分析

註：RIA_1-RIA_20、RIA_36-RIA_45、RIIA_1-RIIA_20、RIIA_36-RIIA_45 所屬分項能力為「整體性的閱讀理解」；RIA_22-RIA_24、RIA_26、RIA_28、RIA_34、RIA_46-RIA_50、RIA_22-RIIA_28、RIIA_30、RIIA_31、RIIA_33、RIIA_35、RIIA_46-RIIA_50 所屬分項能力為「為資訊論證而閱讀」；RIA_21、RIA_25、RIA_27、RIA_29-RIA_33、RIA_35、RIIA_21、RIIA_29、RIIA_32、RIIA_34 所屬分項能力為「導向閱讀」。

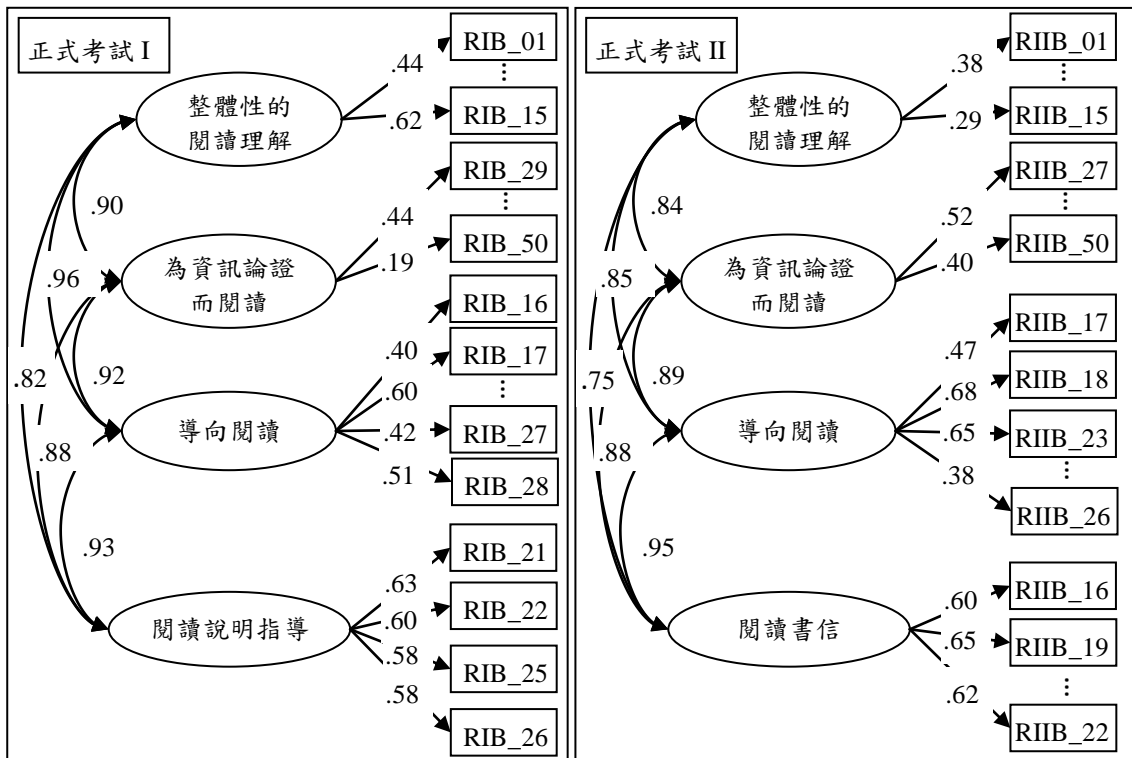


圖 22 進階高階級閱讀測驗四因素驗證性因素分析

註：RIB_1-RIB_15、RIIB_1-RIIB_15 所屬分項能力為「整體性的閱讀理解」；RIB_29-RIB_50、RIIB_27-RIIB_50 所屬分項能力為「為資訊論證而閱讀」；RIB_16-RIB_20、RIB_23、RIB_24、RIB_27、RIB_28、RIB_17、RIB_18、RIB_23、RIB_24 所屬分項能力為「導向閱讀」；RIB_21、RIB_22、RIB_25、RIB_26、RIIB_25、RIIB_26 所屬分項能力為「閱讀說明指導」；RIIB_16、RIIB_19-RIIB_22 所屬分項能力為「閱讀書信」。

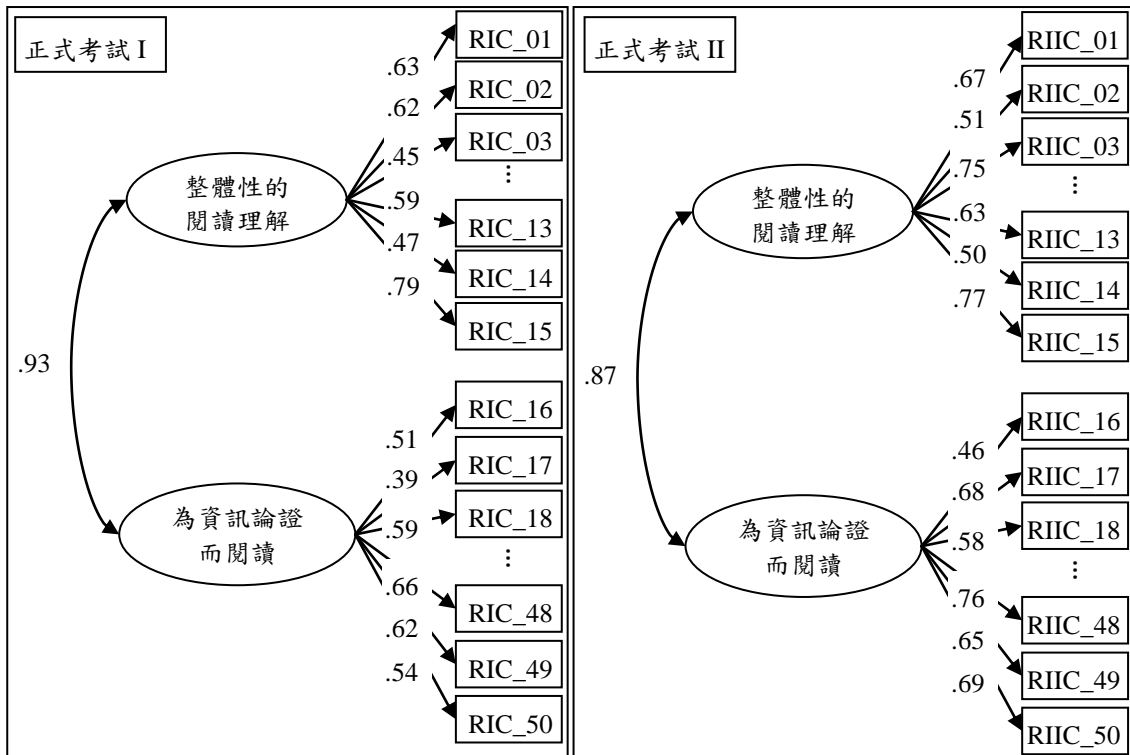


圖 23 流利級閱讀測驗二因素驗證性因素分析

註：RIC_01-RIC_15、RIIC_01-RIIC_15 所屬分項能力為「整體性的閱讀理解」；RIC_16-RIC_50、RIIC_16-RIIC50 所屬分項能力為「為資訊論證而閱讀」。

整體模式適配度主要在評鑑整個模式與觀察資料的適合程度，相當於模式的外在品質。首先，經由卡方自由度比(χ^2/df)評估模式適配度，其原因為卡方分配易受到自由度與估計參數數量的影響，當自由度越大或所需估計的參數個數越多時，卡方值越大，造成假設模式適配度不佳的可能性越大，此時卡方值越不能反映理論模式與觀察資料的適配程度。由於本測驗各等級在進行驗證性因素分析時，自由度均大於 1,000，且模式估計的參數個數均大於 100，故採取卡方自由度比來檢視模式適配度結果。由表 14 可知，閱讀測驗各等級 5 月及 11 月正式考試的卡方自由度比均小於 2，表示模式適配度良好。而在絕對適配度評估上，閱讀測驗各等級 5 月和 11 月正式考試分析結果，RMSEA 數值均小於 .08，表示六種模式皆符合絕對適配度指標。增值適配度評估部分，閱讀測驗各等級 5 月和 11 月正式考試分析結果，CFI 和 NNFI 數值皆大於 .90，顯示六種模式均符合增值適配度指標。

綜合基本適配度與整體適配度之分析結果，可得出以下結論，各等級閱讀測驗具有建構效度，各道試題分別可測得不同的閱讀分項能力，而各閱讀分項能力間的相關係數達到中度至高度相關，反映測得一致之閱讀理解能力。

表 14 整體模式適配度指標摘要表

檢驗模型	正式考試	χ^2/df	RMSEA	CFI	NNFI
入門基礎級二因素模式	I	1.10	.02	.96	.96
	II	1.18	.03	.93	.93
進階高階級三因素模式	I	1.71	.03	.91	.91
	II	1.43	.02	.95	.95
流利級三因素模式	I	1.20	.02	.96	.96
	II	1.12	.02	.98	.97

五、 結論

本文為 2013 年華語文閱讀測驗技術報告，闡述內容共包含兩個部分，第一部分針對華語文閱讀測驗之能力描述、測驗題型與題數、通過門檻等方面分別概述，並說明測驗研發、施測和成績公布之標準化流程。第二部分的內容則以評估 2013 年度華語文閱讀測驗之信度與效度為主，目的為檢視本測驗是否能夠發揮測驗效用，確切地測量受測者的目標潛在能力。

於測驗信度分析方面，本文採內部一致性指標來觀察測驗試題間之相關性，藉以確認整份測驗中的試題是否皆能測量出相同潛在特質、其程度為何。又，在測驗效度分析方面，本測驗首先藉由專家審查、評估各等級試題適切性之步驟來確保試題品質，因具有完整、明確之試題審核步驟，故可作為測驗之內容效度證據。在此應補充說明的是，由於本測驗試題內容的適切性主要仰賴專家審查，且通過審查的試題方能於正式考試使用，故專家審查程序極為重要。為提供更嚴謹、全面的專家效度證據，本會未來將提供外審專家更具針對性的試題檢核項目資料，如設計審題問卷，請專家詳細評分每一道試題，以獲得專家評估試題後的量化結果。

本測驗除了具備測驗之內容效度證據之外，在施測完成後，我們亦針對所收得之受測者作答反應資料，分別進行了試題分析與驗證性因素分析，用以確認受測者之反應資料所建構出的測驗架構，與閱讀測驗研發之初所制訂的目標相同，並以此作為本測驗之建構效度證據。

綜言之，根據本文之信度、效度分析結果顯示，本會於 2013 年度所舉辦之華語文閱讀測驗正式考試，可確實測量到本測驗研發初期所訂定之目標閱讀能力。同時，本測驗所獲得之受測者閱讀能力成績及結果極為可靠，已具備測驗效能。

六、文獻

- 洪小雯(2009)。對外漢語測驗克漏字(完形測驗)文本分析初探。2009 全美中文教師學會年會，美國聖地牙哥。
- 藍珮君、陳柏熹、林玲英、李嘉真、孫雪芹、曾文璇(2013)。Yes/No Angoff 法在華語文聽力測驗與閱讀測驗的應用。中國測驗學會 2013 年會暨心理與教育測驗學術研討會，臺中市、臺灣。
- 藍珮君、陳柏熹(2014)。華語文閱讀測驗信效度分析與垂直等化研究。華語文教學研究 11(1)，99-125。
- Bachman, L. F. (1985). Performance on cloze tests with fixed-ratio and rational deletions. *TESOL Quarterly*, 19(3), 535-555.
- Bagozzi, R.P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16, 74-94.
- Chen, P.-H., Hsu, C.-L., Lan, P.-J., Soon, S. C., Lin, L.-Y., Lee, C.-J., & Tseng, W.-H. (submitted). *Setting standards for multiple-choice items of language tests: methodology and validation – demonstrations of the TOCFL reading and listening tests.*
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests.* Thousand Oaks, CA: Sage.
- Cohen, Andrew D. (1994). *Assessing Language Ability in the Classroom.* Boston, MA.: Heinle and Heinle Publishers.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment* (chap.1 & chap.4). Retrieved January 17, 2007, from http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf
- Dorans, N. J. (2000). Scaling and scales. In Wainer, H (Eds.), *Computerized Adaptive Testing: A Primer* (pp. 135-158). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hu, L.T., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424 – 453.
- Impara, J. C., & Plake, B. S. (1997). Standard-setting: An alternative approach. *Journal of Educational Measurement*, 34, 353–366.
- Kane, M. (1994). Validating the performance standards associated with passing

- scores. *Review of Educational Research*, 64, 425–461.
- Linacre, J. M. (2010). Winsteps® Rasch measurement computer program User's Guide. Beaverton, Oregon: Winsteps.com
- Linacre, J.M. (2009). Winsteps® (Version 3.68.2) [Computer Software]. Beaverton, Oregon: Winsteps.com.
- Muthén, L.K. and Muthén, B.O. (1998-2012). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
- Oller, J. W., Jr. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* Copenhagen: Institute of Educational Research. (Expanded edition, 1980. Chicago: The University of Chicago Press.)

附件 1 閱讀測驗各等級標準設定會議問卷調查結果⁶

附件 1-1 入門基礎級調查結果

問卷題目	平均數
1. 會議帶領者對於本次會議的任務解釋得很清楚。	3.4
2. 會議帶領者對於標準設定方法的操作流程說明得很清楚。	3.7
3. 會議帶領者對於試題難度參數的意涵解釋得很清楚。	3.6
4. 會議帶領者對於 CEFR 的內容和架構說明得很清楚。	3.2
5. 我了解最低能力者在標準設定方法上的涵義。	3.6
6. 最低能力者的表現描述，有助於我判斷最低能力者可以答對的題目。	3.4
7. 標準設定的訓練和練習活動有助於我了解如何進行任務。	3.6
8. 第一回合後團體中的討論和分享，有助於我進行第二回合的判斷。	3.9
9. 第二回合後團體中的討論和分享，有助於我進行第三回合的判斷。	3.9
10. 試題的難度會影響我對試題的判斷。	3.4
11. 我認為自己對每一道試題的判斷，與最低能力者的表現描述一致。	3.3
12. 我對於自己所設定的切截分數(cut score)有信心。	3.4

附件 1-2 進階高階級調查結果

問卷題目	平均數
1. 我認為會議召開前收到的資料，有助於我了解此次標準設定研究的目的。	3.3
2. 會議帶領者對於本次會議的任務解釋得很清楚。	3.8
3. 會議帶領者對於標準設定方法的操作流程說明得很清楚。	3.6
4. 會議帶領者對於 CEFR 的內容和架構說明得很清楚。	4.0
5. 我了解最低能力者在標準設定方法上的涵義。	3.0
6. 最低能力者的表現描述，有助於我判斷最低能力者可以答對的題目。	3.0
7. 標準設定的訓練和練習活動有助於我了解如何進行任務。	3.7
8. 第一回合後團體中的討論和分享，有助於我進行第二回合的判斷。	3.9
9. 第二回合後團體中的討論和分享，有助於我進行第三回合的判斷。	3.6
10. 試題的難度會影響我對試題的判斷。	3.4
11. 我認為自己對每一道試題的判斷，與最低能力者的表現描述一致。	2.9
12. 我對於自己所設定的切截分數(cut score)有信心。	3.0

⁶ 問卷填答方式，「1」表示非常不同意、「2」表示不同意、「3」表示同意、「4」表示非常同意。

附件 1-3 流利精通級調查結果

問卷題目	平均數
1. 我認為會議召開前收到的資料，有助於我了解此次標準設定研究的目的。	2.4
2. 會議帶領者對於本次會議的任務解釋得很清楚。	3.3
3. 會議帶領者對於標準設定方法的操作流程說明得很清楚。	3.4
4. 會議帶領者對於試題難度參數的意涵解釋得很清楚。	3.6
5. 會議帶領者對於 CEFR 的內容和架構說明得很清楚。	2.7
6. 我了解最低能力者在標準設定方法上的涵義。	3.2
7. 最低能力者的表現描述，有助於我判斷最低能力者可以答對的題目。	2.6
8. 標準設定的訓練和練習活動有助於我了解如何進行任務。	3.9
9. 第一回合後團體中的討論和分享，有助於我進行第二回合的判斷。	3.9
10. 第二回合後團體中的討論和分享，有助於我進行第三回合的判斷。	3.9
11. 試題的難度會影響我對試題的判斷。	3.1
12. 我認為自己對每一道試題的判斷，與最低能力者的表現描述一致。	2.9
13. 我對於自己所設定的切截分數(cut score)有信心。	3.0

附件 2 入門基礎級測驗各題因素負荷量及殘差變異量摘要表

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
I	整體性的閱讀理解	RIA_1	.53	0.12	0.00
		RIA_2	.60	0.07	0.00
		RIA_3	.59	0.08	0.00
		RIA_4	.22	0.08	0.00
		RIA_5	.10	0.09	0.29
		RIA_6	.49	0.10	0.00
		RIA_7	.53	0.10	0.00
		RIA_8	.60	0.08	0.00
		RIA_9	.76	0.08	0.00
		RIA_10	.39	0.09	0.00
		RIA_11	.52	0.08	0.00
		RIA_12	.50	0.08	0.00
		RIA_13	.57	0.07	0.00
		RIA_14	.68	0.06	0.00
		RIA_15	.47	0.09	0.00
		RIA_16	.56	0.06	0.00
		RIA_17	.68	0.06	0.00
		RIA_18	.40	0.07	0.00
		RIA_19	.43	0.07	0.00
		RIA_20	.57	0.06	0.00
		RIA_36	.67	0.07	0.00
		RIA_37	.56	0.06	0.00
		RIA_38	.52	0.07	0.00
		RIA_39	.46	0.08	0.00
		RIA_40	.41	0.07	0.00
	RIA_41	.08	0.08	0.32	
	RIA_42	.50	0.08	0.00	
	RIA_43	.74	0.05	0.00	
	RIA_44	.52	0.06	0.00	
	RIA_45	.56	0.06	0.00	
為資訊論證而閱讀	RIA_22	.82	0.08	0.00	
	RIA_23	.84	0.06	0.00	
	RIA_24	.69	0.08	0.00	
	RIA_26	.71	0.09	0.00	
	RIA_28	.35	0.10	0.00	
	RIA_34	.20	0.07	0.01	
	RIA_46	.71	0.05	0.00	
	RIA_47	.83	0.05	0.00	
	RIA_48	.75	0.05	0.00	
	RIA_49	.81	0.04	0.00	
RIA_50	.81	0.05	0.00		
導向閱讀	RIA_21	.19	0.11	0.08	
	RIA_25	.76	0.08	0.00	
	RIA_27	.72	0.07	0.00	
	RIA_29	.67	0.07	0.00	
	RIA_30	.49	0.06	0.00	

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
I	導向閱讀	RIA_31	.65	0.06	0.00
		RIA_32	.67	0.05	0.00
		RIA_33	.57	0.06	0.00
		RIA_35	.57	0.07	0.00
II	整體性的閱讀理解	RIIA_1	.89	0.06	0.00
		RIIA_2	.79	0.07	0.00
		RIIA_3	.73	0.09	0.00
		RIIA_4	.96	0.04	0.00
		RIIA_5	.51	0.11	0.00
		RIIA_6	.78	0.07	0.00
		RIIA_7	.81	0.06	0.00
		RIIA_8	.78	0.07	0.00
		RIIA_9	.77	0.07	0.00
		RIIA_10	.72	0.06	0.00
		RIIA_11	.27	0.10	0.01
		RIIA_12	.74	0.08	0.00
		RIIA_13	.13	0.10	0.19
		RIIA_14	.54	0.08	0.00
		RIIA_15	.30	0.09	0.00
		RIIA_16	.69	0.07	0.00
		RIIA_17	.76	0.06	0.00
		RIIA_18	.67	0.07	0.00
		RIIA_19	.74	0.07	0.00
		RIIA_20	.48	0.08	0.00
		RIIA_36	.46	0.07	0.00
		RIIA_37	.72	0.08	0.00
		RIIA_38	.66	0.08	0.00
		RIIA_39	.24	0.10	0.01
		RIIA_40	.33	0.09	0.00
		RIIA_41	.52	0.11	0.00
	RIIA_42	-.07	0.07	0.35	
	RIIA_43	.65	0.08	0.00	
	RIIA_44	.64	0.07	0.00	
	RIIA_45	.54	0.09	0.00	
	為資訊論證而閱讀	RIIA_22	.65	0.11	0.00
		RIIA_23	.72	0.08	0.00
		RIIA_24	.75	0.11	0.00
RIIA_25		.83	0.06	0.00	
RIIA_26		.68	0.09	0.00	
RIIA_27		.76	0.09	0.00	
RIIA_28		.70	0.08	0.00	
RIIA_30		.61	0.09	0.00	
RIIA_31		.82	0.07	0.00	
RIIA_33		.67	0.06	0.00	
RIIA_35		.82	0.04	0.00	
RIIA_46		.91	0.05	0.00	
RIIA_47	.78	0.04	0.00		

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
II	為資訊論證而閱讀	RIIA_48	.66	0.07	0.00
		RIIA_49	.72	0.06	0.00
		RIIA_50	.71	0.06	0.00
	導向閱讀	RIIA_21	.12	0.12	0.32
		RIIA_29	.65	0.12	0.00
		RIIA_32	.72	0.09	0.00
		RIIA_34	.58	0.08	0.00

附件 3 進階高階級測驗各題因素負荷量及殘差變異量摘要表

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
I	整體性的閱讀理解	RIB_1	.44	0.05	0.00
		RIB_2	.40	0.05	0.00
		RIB_3	.27	0.04	0.00
		RIB_4	.37	0.05	0.00
		RIB_5	.52	0.05	0.00
		RIB_6	.42	0.04	0.00
		RIB_7	.35	0.04	0.00
		RIB_8	.12	0.05	0.01
		RIB_9	.60	0.04	0.00
		RIB_10	.59	0.04	0.00
		RIB_11	.55	0.04	0.00
		RIB_12	.70	0.03	0.00
		RIB_13	.43	0.05	0.00
		RIB_14	.47	0.04	0.00
		RIB_15	.62	0.04	0.00
	為資訊論證而閱讀	RIB_29	.44	0.04	0.00
		RIB_30	.53	0.04	0.00
		RIB_31	.45	0.04	0.00
		RIB_32	.42	0.04	0.00
		RIB_33	.57	0.04	0.00
		RIB_34	.67	0.03	0.00
		RIB_35	.58	0.04	0.00
		RIB_36	.45	0.04	0.00
		RIB_37	.38	0.04	0.00
		RIB_38	.43	0.04	0.00
		RIB_39	.67	0.03	0.00
		RIB_40	.46	0.04	0.00
		RIB_41	.44	0.05	0.00
		RIB_42	.21	0.05	0.00
		RIB_43	.37	0.04	0.00
RIB_44	.33	0.05	0.00		
RIB_45	.12	0.06	0.03		
RIB_46	.22	0.05	0.00		
RIB_47	.22	0.06	0.00		
RIB_48	.52	0.04	0.00		
RIB_49	.40	0.05	0.00		
RIB_50	.19	0.05	0.00		
導向閱讀	RIB_16	.40	0.04	0.00	
	RIB_17	.60	0.04	0.00	
	RIB_18	.54	0.04	0.00	
	RIB_19	.54	0.04	0.00	
	RIB_20	.62	0.04	0.00	
	RIB_23	.59	0.04	0.00	
	RIB_24	.63	0.04	0.00	
	RIB_27	.42	0.04	0.00	
RIB_28	.51	0.04	0.00		

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
I	閱讀說明指導	RIB_21	.63	0.04	0.00
		RIB_22	.60	0.04	0.00
		RIB_25	.58	0.04	0.00
		RIB_26	.58	0.04	0.00
II	整體性的閱讀理解	RIIB_1	.38	0.05	0.00
		RIIB_2	.23	0.05	0.00
		RIIB_3	.78	0.03	0.00
		RIIB_4	.39	0.05	0.00
		RIIB_5	.55	0.05	0.00
		RIIB_6	.50	0.04	0.00
		RIIB_7	.57	0.04	0.00
		RIIB_8	.71	0.04	0.00
		RIIB_9	.22	0.06	0.00
		RIIB_10	.56	0.04	0.00
		RIIB_11	.15	0.06	0.02
		RIIB_12	.49	0.05	0.00
		RIIB_13	.52	0.05	0.00
		RIIB_14	.19	0.05	0.00
		RIIB_15	.29	0.05	0.00
	為資訊論證而閱讀	RIIB_27	.52	0.05	0.00
		RIIB_28	.42	0.05	0.00
		RIIB_29	.47	0.05	0.00
		RIIB_30	.42	0.05	0.00
		RIIB_31	.59	0.04	0.00
		RIIB_32	.85	0.03	0.00
		RIIB_33	.61	0.04	0.00
		RIIB_34	.72	0.04	0.00
		RIIB_35	.59	0.04	0.00
		RIIB_36	.59	0.04	0.00
		RIIB_37	.66	0.04	0.00
		RIIB_38	.47	0.04	0.00
		RIIB_39	.50	0.04	0.00
		RIIB_40	.37	0.05	0.00
		RIIB_41	.68	0.04	0.00
		RIIB_42	.59	0.04	0.00
		RIIB_43	.72	0.03	0.00
		RIIB_44	.63	0.04	0.00
		RIIB_45	.33	0.05	0.00
		RIIB_46	.63	0.04	0.00
		RIIB_47	.52	0.04	0.00
		RIIB_48	.44	0.05	0.00
		RIIB_49	.49	0.04	0.00
	RIIB_50	.40	0.05	0.00	
	導向閱讀	RIIB_17	.47	0.05	0.00
		RIIB_18	.68	0.04	0.00
		RIIB_23	.65	0.04	0.00
		RIIB_24	.68	0.04	0.00

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
II	導向閱讀	RIIB_25	.27	0.06	0.00
		RIIB_26	.38	0.05	0.00
	閱讀書信	RIIB_16	.60	0.04	0.00
		RIIB_19	.65	0.05	0.00
		RIIB_20	.74	0.04	0.00
		RIIB_21	.58	0.05	0.00
		RIIB_22	.62	0.05	0.00

附件 4 流利級測驗各題因素負荷量及殘差變異量摘要表

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
I	整體性的閱讀理解	RIC_1	.63	0.05	0.00
		RIC_2	.62	0.05	0.00
		RIC_3	.45	0.06	0.00
		RIC_4	.50	0.06	0.00
		RIC_5	.67	0.05	0.00
		RIC_6	.41	0.06	0.00
		RIC_7	.60	0.05	0.00
		RIC_8	.51	0.06	0.00
		RIC_9	.68	0.05	0.00
		RIC_10	.64	0.05	0.00
		RIC_11	.34	0.07	0.00
		RIC_12	.62	0.06	0.00
		RIC_13	.59	0.05	0.00
		RIC_14	.47	0.06	0.00
		RIC_15	.79	0.05	0.00
	為資訊論證而閱讀	RIC_16	.51	0.07	0.00
		RIC_17	.39	0.06	0.00
		RIC_18	.59	0.05	0.00
		RIC_19	.72	0.04	0.00
		RIC_20	.52	0.05	0.00
		RIC_21	.50	0.06	0.00
		RIC_22	.61	0.05	0.00
		RIC_23	.37	0.06	0.00
		RIC_24	.34	0.07	0.00
		RIC_25	.39	0.06	0.00
		RIC_26	.58	0.05	0.00
		RIC_27	.68	0.04	0.00
		RIC_28	.48	0.06	0.00
		RIC_29	.72	0.04	0.00
		RIC_30	.49	0.06	0.00
		RIC_31	.54	0.05	0.00
		RIC_32	.48	0.06	0.00
		RIC_33	.56	0.05	0.00
		RIC_34	.55	0.06	0.00
		RIC_35	.33	0.07	0.00
		RIC_36	.51	0.06	0.00
		RIC_37	.28	0.08	0.00
		RIC_38	.60	0.05	0.00
		RIC_39	.64	0.05	0.00
		RIC_40	.68	0.05	0.00
		RIC_41	.46	0.06	0.00
		RIC_42	.71	0.05	0.00
		RIC_43	.71	0.05	0.00
		RIC_44	.62	0.05	0.00
		RIC_45	.62	0.07	0.00
		RIC_46	.47	0.07	0.00

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
I	為資訊論證而閱讀	RIC_47	.54	0.05	0.00
		RIC_48	.66	0.05	0.00
		RIC_49	.62	0.06	0.00
		RIC_50	.54	0.06	0.00
II	整體性的閱讀理解	RIIC_1	.67	0.06	0.00
		RIIC_2	.51	0.07	0.00
		RIIC_3	.75	0.05	0.00
		RIIC_4	.37	0.08	0.00
		RIIC_5	.67	0.06	0.00
		RIIC_6	.58	0.07	0.00
		RIIC_7	.69	0.06	0.00
		RIIC_8	.45	0.07	0.00
		RIIC_9	.61	0.06	0.00
		RIIC_10	.54	0.06	0.00
		RIIC_11	.34	0.08	0.00
		RIIC_12	.44	0.07	0.00
		RIIC_13	.63	0.06	0.00
		RIIC_14	.50	0.06	0.00
		RIIC_15	.77	0.06	0.00
	為資訊論證而閱讀	RIIC_16	.46	0.07	0.00
		RIIC_17	.68	0.07	0.00
		RIIC_18	.58	0.07	0.00
		RIIC_19	.58	0.06	0.00
		RIIC_20	.59	0.06	0.00
		RIIC_21	.46	0.07	0.00
		RIIC_22	.56	0.06	0.00
		RIIC_23	.55	0.06	0.00
		RIIC_24	.65	0.05	0.00
		RIIC_25	.67	0.05	0.00
		RIIC_26	.50	0.07	0.00
		RIIC_27	.68	0.05	0.00
		RIIC_28	.64	0.05	0.00
		RIIC_29	.67	0.05	0.00
		RIIC_30	.61	0.06	0.00
		RIIC_31	.72	0.05	0.00
		RIIC_32	.64	0.05	0.00
		RIIC_33	.68	0.05	0.00
	RIIC_34	.51	0.07	0.00	
	RIIC_35	.68	0.05	0.00	
	RIIC_36	.64	0.05	0.00	
	RIIC_37	.49	0.06	0.00	
	RIIC_38	.60	0.06	0.00	
	RIIC_39	.71	0.05	0.00	
	RIIC_40	.62	0.07	0.00	
	RIIC_41	.65	0.05	0.00	
	RIIC_42	.57	0.06	0.00	
	RIIC_43	.57	0.06	0.00	

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
II	為資訊論證而閱讀	RIIC_44	.72	0.05	0.00
		RIIC_45	.52	0.07	0.00
		RIIC_46	.67	0.05	0.00
		RIIC_47	.34	0.08	0.00
		RIIC_48	.76	0.04	0.00
		RIIC_49	.65	0.05	0.00
		RIIC_50	.69	0.05	0.00

書名：華語文能力測驗技術報告—2013(2)
閱讀測驗信效度

出版者：國家華語測驗推動工作委員會
24449 新北市林口區仁愛路一段 2 號
886-2-7734-5638

印刷者：上校文化印刷有限公司
80744 高雄市三民區通化街 88 巷 26 號
886-7-311-6011

出版日期：2015 年 12 月

定價：新台幣 100 元

版權所有

翻印必究