

# 華語文能力測驗技術報告 — 2014(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布之標準化流程，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.....	2
(三) 通過門檻.....	4
三、 測驗標準化流程.....	5
(一) 試題收集.....	5
(二) 試題修審.....	6
(三) 預試.....	6
(四) 試題分析.....	7
(五) 輸入題庫.....	8
(六) 組合正式卷.....	8
(七) 檢核正式卷與多媒體檔案.....	9
(八) 成績公布.....	9
四、 測驗評估.....	10
(一) 信度.....	10
1. 內部一致性信度.....	11
2. 測驗訊息量.....	11
(二) 效度.....	15
1. 內容效度.....	15
2. 建構效度.....	16
五、 結論.....	32
六、 文獻.....	34



## 表目錄

表 1	基本能力描述 .....	2
表 2	測驗題型 .....	3
表 3	測驗通過分數 .....	4
表 4	各等級正式考試卷難度分布 .....	9
表 5	內部一致性信度摘要表 .....	11
表 6	雙向細目表 .....	16
表 7	試題難度估計分布 .....	17
表 8	受測者能力估計分布 .....	17
表 9	試題適配分布 .....	20
表 10	入門基礎級測驗不適配試題選項描述性摘要表 .....	21
表 11	入門基礎級不適配試題之預試適配分析結果.....	22
表 12	進階高階級測驗不適配試題選項描述性摘要表 .....	23
表 13	進階高階級不適配試題之預試適配分析結果 .....	24
表 14	流利精通級測驗不適配試題選項描述性摘要表 .....	25
表 15	整體模式適配度指標摘要及模式比較表 .....	31





## 圖目錄

圖 1	正式考試標準化流程 .....	5
圖 2	入門基礎級測驗訊息量 .....	13
圖 3	入門基礎級測驗測量誤差 .....	13
圖 4	進階高階級測驗訊息量 .....	13
圖 5	進階高階級測驗測量誤差 .....	14
圖 6	流利精通級測驗訊息量 .....	14
圖 7	流利精通級測驗測量誤差 .....	14
圖 8	入門基礎級測驗受測者與試題分布圖 .....	18
圖 9	進階高階級測驗受測者與試題分布圖 .....	18
圖 10	流利精通級測驗受測者與試題分布圖 .....	19
圖 11	入門基礎級閱讀測驗驗證性因素分析 .....	28
圖 12	進階高階級閱讀測驗驗證性因素分析 .....	29
圖 13	流利精通級閱讀測驗驗證性因素分析 .....	30



## 附件目錄

附件 1	入門基礎級測驗各題因素負荷量及殘差變異量摘要表.....	36
附件 2	進階高階級測驗各題因素負荷量及殘差變異量摘要表.....	38
附件 3	流利精通級測驗各題因素負荷量及殘差變異量摘要表.....	40



## 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，並於 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在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。

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、進階高階級與流利精通級，各等再依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，共六級。應試者可考量自身學習背景或語言能力選擇合適的等級應考。相較於僅能判別受測者通過與否的測驗，此一分等架構可更進一步區分出通過測驗者能力的落點；同時，對於應試者及試務工作者來說，將更符合經濟效益。例如，改版後的測驗方式(各等分為兩級)，應試者只需參加一次測驗，即可同時判斷其在兩個等級內之能力水平。

本報告首先針對 2014 年華語文閱讀測驗各等之能力描述、測驗題型、通過門檻、測驗實施與成績公布之標準化流程進行概述。其次，分別闡述本年度正式考試之信、效度分析結果。最後，則將根據各項分析結果，提出相關討論意見。

## 二、 簡介

2014 年本測驗正式考試含三等六級：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)；流利精通級(Band C)，區分為流利級(Level 5)與精通級(Level 6)。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)、C1(Effective Operational Proficiency)與 C2(Mastery)。以下分就各等之能力描述、測驗題型、通過門檻三方面進行介紹。

### (一) 能力描述

各等級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「日常生活的一般簡易溝通能力」；進階高階級以「語言段落的理解分析能力」為主；流利精通級則著重於「語言使用的廣度與精熟度」。

表 1 基本能力描述

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	在閱讀時具有相當大的自主性，能因應不同的文本及目的，採用不同的閱讀策略和速度。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。
精通級	能理解幾乎所有書寫語言的形式，包括抽象、結構複雜或者相當口語化的文學或非文學作品等。

### (二) 測驗題型

本測驗以各等級語言學習者之能力表現(如表 1)及所預設之測知目標為參考依據，在入門基礎級測驗中採行「單句理解」、「看圖釋義」、「選詞填空」與「完

成段落」四個題型；進階高階級與流利精通級則均採用「選詞填空」與「閱讀理解」兩個題型，各等級題型如表 2 所示。入門基礎級前三部分為三選一之單選題，第四部分為配合題；在進階高階級與流利精通級中兩部分題型皆為四選一之單選題。一卷總題數皆為 50 題，測驗時間為 60 分鐘。

表 2 測驗題型

等級/題型	單句理解	看圖釋義	完成段落	選詞填空	閱讀理解
入門基礎級	✓	✓	✓	✓	--
進階高階級	--	--	--	✓	✓
流利精通級	--	--	--	✓	✓

由表 2 可知，各等級皆包含了「選詞填空」題型，其原因在於，「選詞填空」之題型設計主要是為了測得受測者所掌握的語言知識。在外語學習的過程中，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」題型在各等測驗中均被採用。本測驗之「選詞填空」題型乃是採用多項選擇式克漏字測驗(Multiple-choice Cloze Test)搭配意向刪詞法(rational deletion)設計而成，目的在設計出不同體裁、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，在評量學習者的閱讀表現時，文本語言的複雜度、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(特別是當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)。由於入門基礎級受測者的語言能力尚處於初級階段，為避免其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，多採用情境圖片作為試題文字的輔助；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；而達到流利精通級水準的受測者，已不需要額外輔助即可了解文本內容(Council of Europe, 2001)，因此皆未設計圖文輔助一類題型。

### (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; 國家華語測驗推動工作委員會, 2015)。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之標準化流程執行。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均獲得支持，驗證華語文閱讀測驗將華語學習者依其報考等級與測驗表現區分為 A1 至 C2 六個等級的有效性。標準設定程序各步驟說明及詳細內容，可參考藍珮君等人(2013)與國家華語測驗推動工作委員會(2015)。

自 2013 年起，本測驗正式考試受測者的測驗結果與通過等級之判定依據，改以量尺分數(scale score)呈現。量尺分數是將考生能力值透過制定的平均數與標準差進行轉換而得到的測驗分數，其目的在使測驗通過標準能不受各次組卷試題間些微難度變化的影響，以確保測驗的客觀、公平。本測驗之量尺分數乃由測驗專家根據以往參與本會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平均數皆為 50，標準差均為 10，測驗滿分均為 80 分，各等級通過分數如表 3 所示。

表 3 測驗通過分數

測驗等級	通過等級	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
流利精通級	流利級	52
	精通級	69



### 三、 測驗標準化流程

2014 年本測驗標準化流程(如圖 1 所示)主要可分為八個步驟，依序為試題收集、試題修審、預試、試題分析、題庫輸入、組合正式卷、檢核正式卷與多媒體檔案，以及成績公布。茲說明如下：

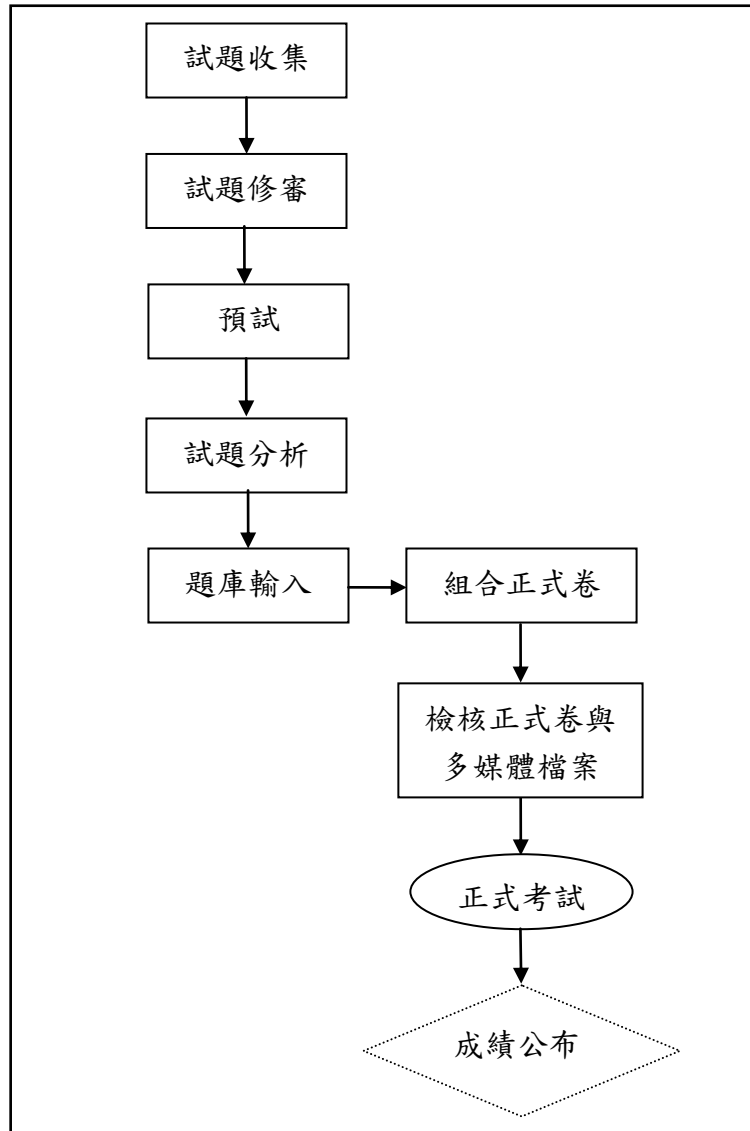


圖 1 正式考試標準化流程

#### (一) 試題收集

2014 年協助閱讀測驗各等級命題工作之教師共計 34 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、提出修改意見，並與命題教師討論修改內容。待確定收題後，

再將需繪製圖片的試題交由專業繪圖人員繪製，繪圖時間約為三週。本年度命題回收數量共計 818 道試題，分別為入門基礎級 367 題、進階高階級 207 題、流利精通級 244 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，待逐一審查每道試題編寫是否合宜、試卷難度分布是否合理等項目後，提出審題意見，審題時間約為三週。研發人員隨後根據會內初審意見再次修改試題。如此，方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，逐題審查文本預設等級及試題內容是否符合該等級應有之標準。如：圈選文本預設等級評估量表；審查試題設計是否確實測知考生相應的閱讀能力；評估文本主題選擇、提問及選項設計之適切性。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**

第二階段專家學者外審修改後之試題，將由本會執行秘書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，再由研發人員根據審題意見修改試題。

### **4. 製作試題相關媒體檔案**

本階段乃針對上述審題三步驟完成之定稿，進行試題送繪與繪圖定稿等工作。

## **(三) 預試**

修審後的試題皆須進行預試，並完成樣本收集程序，以利通過量化數據客觀地評估試題品質。由於預試樣本來自世界各地，無法進行分層隨機抽樣，取樣方式為便利取樣。收樣時為使樣本具有代表性，盡量控制每一份預試卷的樣本涵蓋

亞洲、美洲、歐洲地區考生。此外，亦避免同一預試卷樣本只來自單一地區。本年度預試規劃之總施測題數分別為入門基礎級 200 題、進階高階級 250 題及流利精通級 200 題。全年度之預試樣本分別來自臺灣地區及海外地區，受測樣本人數分別為入門基礎級 1,777 名、進階高階級 1,183 名及流利精通級 1,829 名，海外地區的考生約占了 21%。

#### (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(item response theory；簡稱 IRT)為分析取向。由於本測驗中各試題之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(Rasch, 1960)進行資料分析，Rasch 模式如公式(1)所示。

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

由公式(1)可發現，Rasch 模式的假設為，受測者答對每一道試題的機率受到受測者具備的能力及試題難度的影響，故將受測者在各試題上的反應透過公式(1)進行計算，即可估計出受測者的能力及試題的難易度，並使應考不同試題的受測者之能力可以互相比較。由於在 IRT 中，受測者的能力參數與試題參數被同時納入相同模式裡，因此，在估計受測者能力時，已考量了試題參數的影響；而在估計試題參數時，也考量了受測者能力的影響。

本測驗以 Winsteps 3.68.2 版(Linacre, 2009)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(mean-square)，及其標準化 Z 值(z-standardized)，兩者分別簡稱為 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，本測驗亦提供試題鑑別度指標<sup>1</sup>、難度 P 值<sup>2</sup>及選項分析等資料作為輔助，以期透過更全面的觀點評估試題品質。

---

<sup>1</sup> 試題鑑別度指標為 D 值及點二系列相關係數。

<sup>2</sup> 難度 P 值為高低分組受測者答對率之平均數。

## (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研討，以決定試題是否予以保留。扣除共同題後，本年度輸入閱讀測驗題庫之試題數量分別為入門基礎級 207 題、進階高階級 172 題及流利精通級 220 題，共計 599 題<sup>3</sup>。

## (六) 組合正式卷

閱讀測驗正式卷乃由題庫內之試題所組成，各卷所挑選之試題，其難度參數均符合試卷難度區間要求，各難度區間之題數亦依照固定比例分配。組卷題數分布基本原則為切截點附近各配置 24% 試題，以確保通過門檻能力值的估計精準度；低於切截點的試題占 40%，降低程度較低考生作答時的挫折感；高於切截點的試題占 12%，以區辨程度較高考生。

華語文閱讀測驗自 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後（藍珮君、陳柏熹，2014），試題難度量尺即為同一量尺，各測驗等級為因應各等級的通過門檻不同，而在組卷時有不同的主要難度區間範圍，依序分別為：入門基礎級介於-3.99 至-0.50 logits 之間；進階高階級介於-1.49 至 2.50 logits 之間；流利精通級介於 1.00 至 4.00 logits 之間。

正式卷初稿完成後，即由研發人員檢核試題之主題分布是否過度集中、考點是否重複、試題難度是否符合雙向細目表之規劃等，並提出建議，再根據檢核建議更換試題，直至整份試卷皆符合上述要求後，始得定稿。2014 年各等級國內正式考試卷之難度分布如表 4 所示。

---

<sup>3</sup> 包含前一年度開始預試，跨年度於今年完成收樣之試題。

表 4 各等級正式考試卷難度分布

試題難度	入門基礎級		進階高階級		流利精通級	
	I	II	I	II	I	II
<-3.99	2%	0%	-	-	-	-
-3.99~-3.50	16%	14%	-	-	-	--
-3.49~-3.00	14%	14%	-	-	-	-
-2.99~-2.50	24%	30%	2%	-	-	-
-2.49~-2.00	10%	4%	-	2%	-	-
-1.99~-1.50	14%	10%	4%	4%	-	-
-1.49~-1.00	12%	12%	16%	10%	-	-
-0.99~-0.50	8%	16%	24%	20%	-	-
-0.49~0.00	-	-	8%	14%	-	-
0.01~0.50	-	-	8%	10%	2%	4%
0.51~1.00	-	-	8%	10%	6%	6%
1.01~1.50	-	-	10%	10%	18%	24%
1.51~2.00	-	-	12%	14%	30%	26%
2.01~2.50	-	-	8%	6%	24%	16%
2.51~3.00	-	-	-	-	10%	14%
3.01~3.50	-	-	-	-	6%	6%
>3.50	-	-	-	-	4%	4%

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，此一步驟除針對同一試卷考點不重複、排除試題互斥、確認題幹主題多元性等項目進行複檢之外，還將針對所輸入試題之文字及圖片信息之清晰度、正確性等方面進行檢查。待確認無誤後，研發人員即可轉出卷包，並將試卷派送至考試系統端進行模擬測試。模擬測試時，先由研發人員進行首輪測試，隨後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。

### (八) 成績公布

正式考試結束後，待檢核並確認受測者作答資料無誤，則進行成績報表檢核，亦經確認無誤後，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 2014 年華語文閱讀測驗正式考試之信效度，來說明本年度之測驗效能。2014 年度本測驗正式考試分別於 5 月 3 日、11 月 1 日舉行。五月份之到考人數分別為：入門基礎級 433 名、進階高階級 866 名及流利精通級 338 名。十一月份到考人數則分別為：入門基礎級 329 名、進階高階級 727 名及流利精通級 305 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。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的穩定性；
3. 內部一致性信度，表示測驗所測量到的潛在特質的一致性；
4. 評分者信度，表示經由不同評分者所得到的評分結果間的一致性。

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，是以測驗訊息量(test information： $TI(\theta)$ )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  $\alpha$  係數及庫李 (Kuder-Richardson) 20 號或 21 號公式。本節將分析 2014 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、進階高階級與流利精通級測驗兩次正式考試之內部一致性信度如表 5 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 .85 至 .91 之間；在第二次正式考試中，各等級測驗信度係數介於 .89 至 .95 之間。由此可知，2014 年華語文閱讀測驗正式考試卷信度良好。

表 5 內部一致性信度摘要表

正式考試	入門基礎級	進階高階級	流利精通級
I	.91	.85	.89
II	.95	.89	.92

## 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  $\theta$  的受測者所提供的訊息量； $P_i'(\theta)$  為  $P_i(\theta)$  對  $\theta$  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  $\theta$  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； $\sigma$  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、進階高階級和流利精通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 44 分時最大，數值為 10.28，測量誤差為 3.12，而在入門級與基礎級門檻的測驗訊息量分別為 10.19 與 7.37，測量誤差分別為 3.13 與 3.68，信度係數分別為.95 與.93；第二次正式考試之測驗訊息量以量尺分數 44、46 分為最大，數值均為 10.06，測量誤差均為 3.15，在入門級與基礎級門檻的測驗訊息量分別為 9.91 與 7.73，測量誤差分別為 3.18 與 3.60，信度係數分別為.96 與.95。上述結果顯示，入門基礎級測驗在兩個通過門檻所提供的測量訊息，相較於其他分數而言，具有較高的訊息量與較低的測量誤差，且兩個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精通級測驗(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.85 至.94 之間，具有良好信度。如進階高階級測驗之兩個通過門檻在兩次正式考試的測驗訊息量各為 9.69 與 7.65，以及 9.40 及 7.83，測量誤差分別為 3.21 與 3.62，及 3.26 與 3.57，信度係數為.90 與.87，及.91 與.89。而流利精通級測驗兩個門檻在兩次正式考試的測驗訊息量則分別為 10.97 與 5.89，及 10.73 與 5.83，測量誤差為 3.02 與 4.12，及 3.05 與 4.14，信度係數分別為.92 與.85，及.94 與.89。



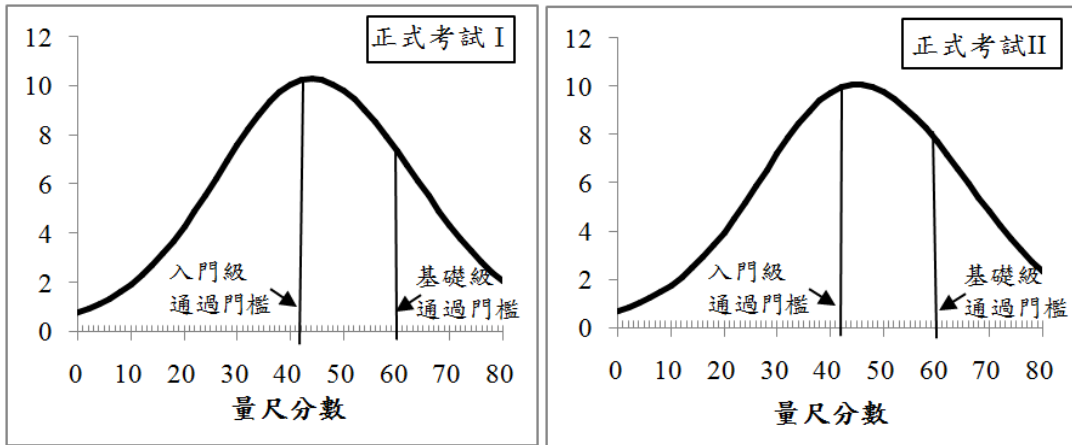


圖 2 入門基礎級測驗訊息量

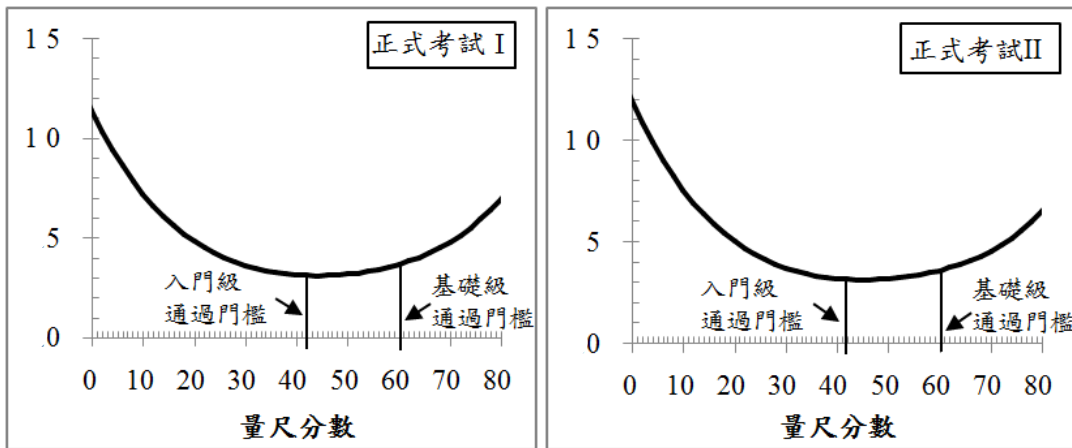


圖 3 入門基礎級測驗測量誤差

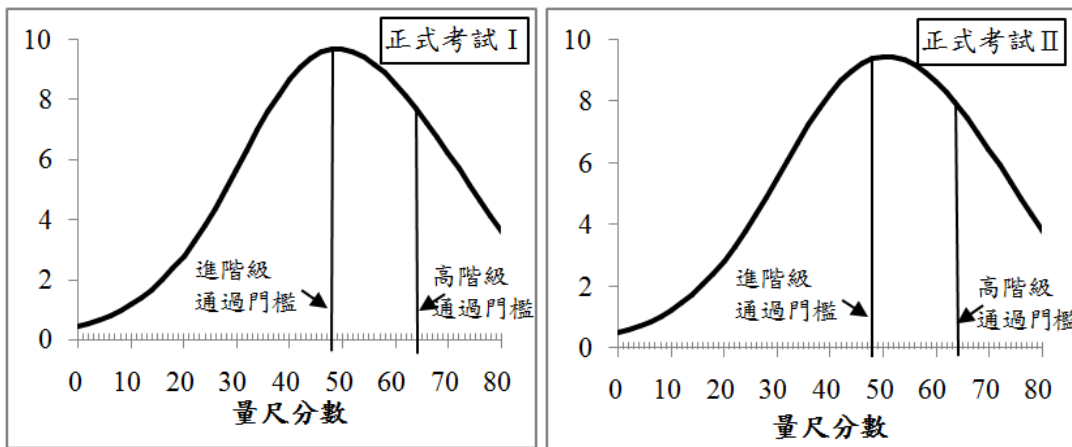


圖 4 進階高階級測驗訊息量

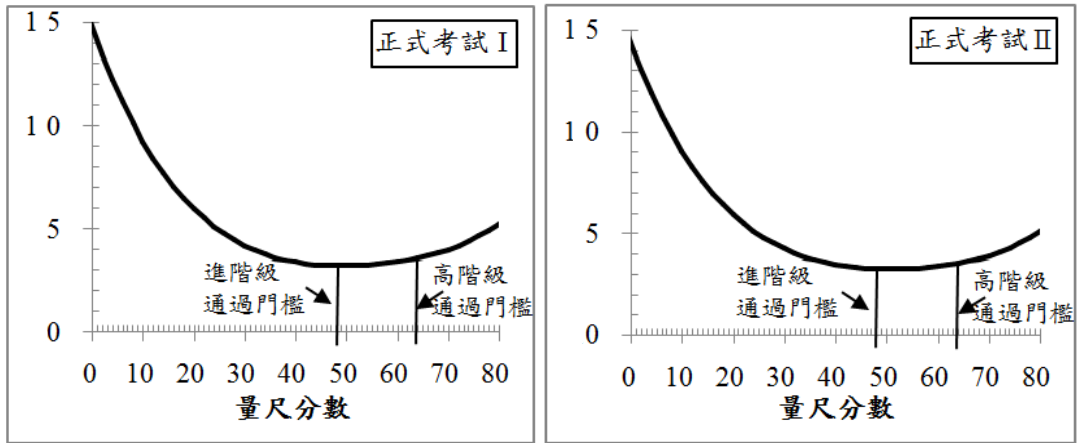


圖 5 進階高階級測驗測量誤差

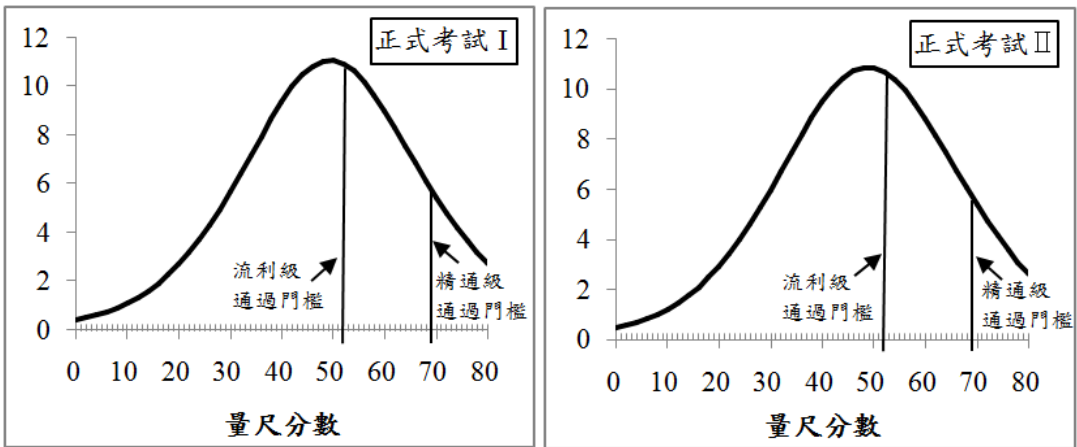


圖 6 流利精通級測驗訊息量

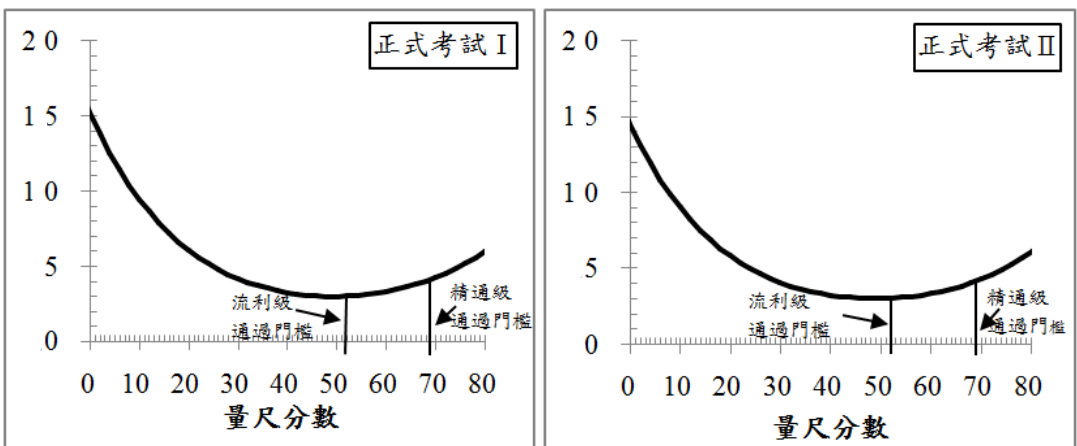


圖 7 流利精通級測驗測量誤差

## (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。常作為測驗效度證據者主要可分為三大類，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 2014 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 6)。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(包含題幹、選項及提問描述)詞彙使用適切性和表達適當性。若出現不符原則者，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標準之試題，進行再檢視與再修改，最後將試題組成預試卷，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 6 雙向細目表

分項能力	題型 測驗等級	單句	看圖	完成	選詞填空			材料	短文閱讀	
		理解	釋義	段落	入門	進階	流利	閱讀	進階	流利
		入門 基礎	入門 基礎	入門 基礎	入門 基礎	高階	精通	高階	高階	精通
整體性的閱讀理解		V		V	V	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					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，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。

在試題難度分布上，整體而言，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 0.14 logit；標準差的絕對值差異小於 0.12 (如表 7)。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 4)。

在考生能力分布方面(參見表 8)，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 0.38 logit；各等級考生平均能力估計標準差的絕對值差異均小於 0.32。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-0.38 與-0.44，50 題試題的平均難度估計值分別為-2.48 及-2.35。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 2.10 及 1.91 logits，相對於受測者能力而言，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 45% 與 47%。

在進階高階級與流利精通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入門基礎級小，乃介於 0.05 至 0.69 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配範圍。透過上述結果可知，測驗等級較高時，考生平均能力估計值與試題平均難度估計值差異較小。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 93%；流利精通級則分別為 91% 與 81%。

**表 7 試題難度估計分布**

測驗等級	入門基礎級		進階高階級		流利精通級	
	I	II	I	II	I	II
正式考試						
平均值	-2.48	-2.35	0.06	0.20	2.02	1.93
標準差	0.97	1.01	1.31	1.19	0.85	0.83
最大值	-0.54	-0.60	2.19	2.37	5.19	4.06
最小值	-4.01	-3.82	-2.79	-2.16	0.23	0.21

**表 8 受測者能力估計分布**

測驗等級	入門基礎級		進階高階級		流利精通級	
	I	II	I	II	I	II
正式考試						
平均值	-0.38	-0.44	0.51	0.89	2.07	2.35
標準差	1.37	1.69	1.02	1.10	1.05	1.26
最大值	3.10	3.25	5.94	4.74	5.13	6.17
最小值	-3.99	-4.32	-5.77	-1.88	-3.38	-1.53

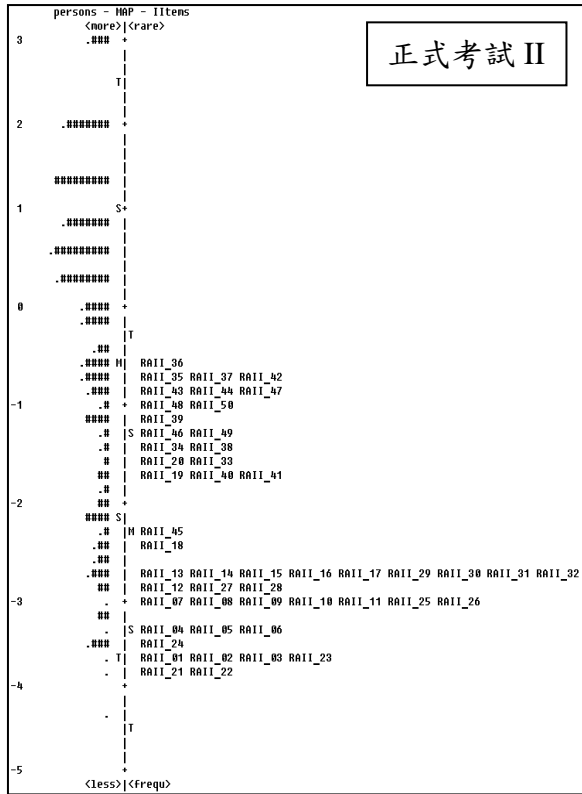
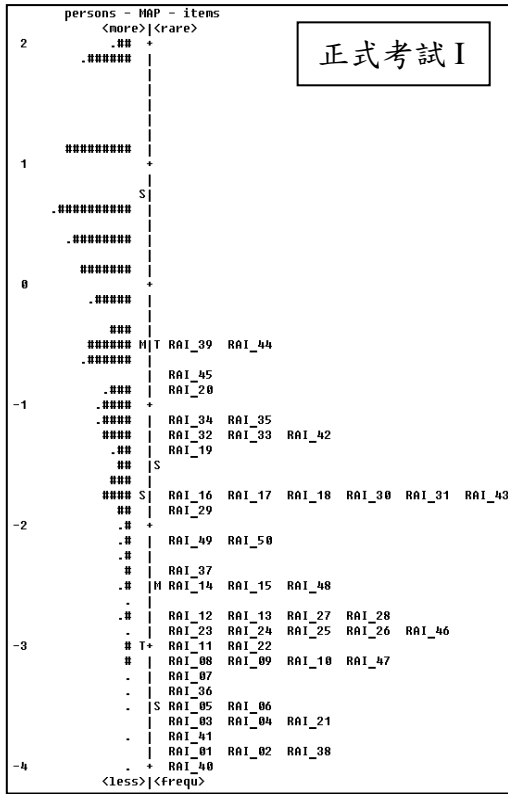


圖 8 入門基礎級測驗受測者與試題分布圖

註：“#”在正式考試 I 與 II 分別表示四位及三位受測者；“.”表示一位受測者；RAI\_01~RAI\_50 和 RAI1\_01~RAI1\_50 表示試題編號。

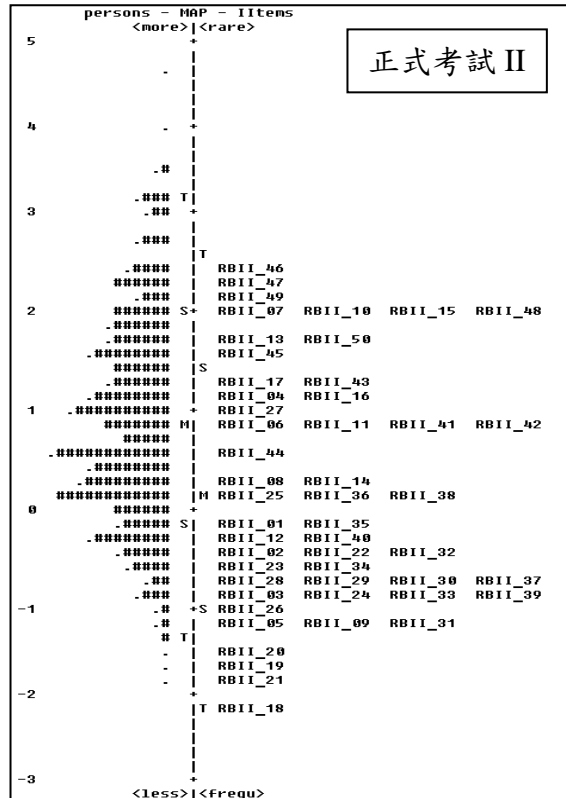
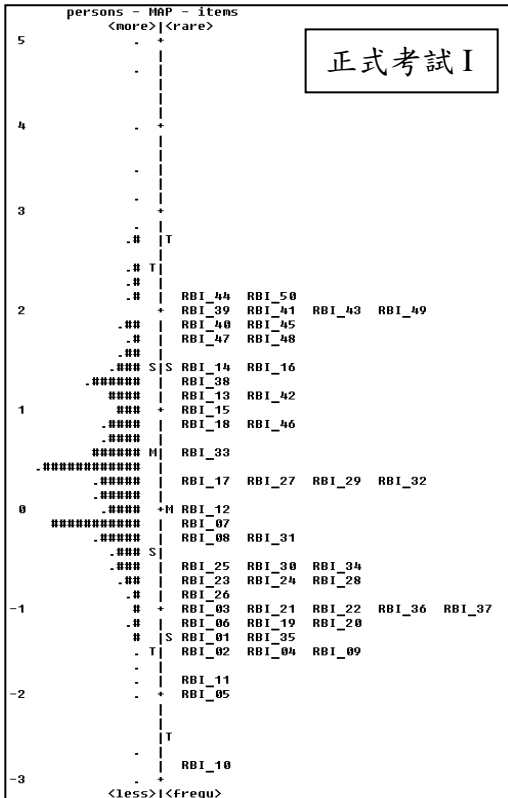


圖 9 進階高階級測驗受測者與試題分布圖

註：“#”在正式考試 I 與 II 分別表示八位及四位受測者；“.”表示一位受測者；RBI\_01~RBI\_50 和 RBII\_01~RBII\_50 表示試題編號。

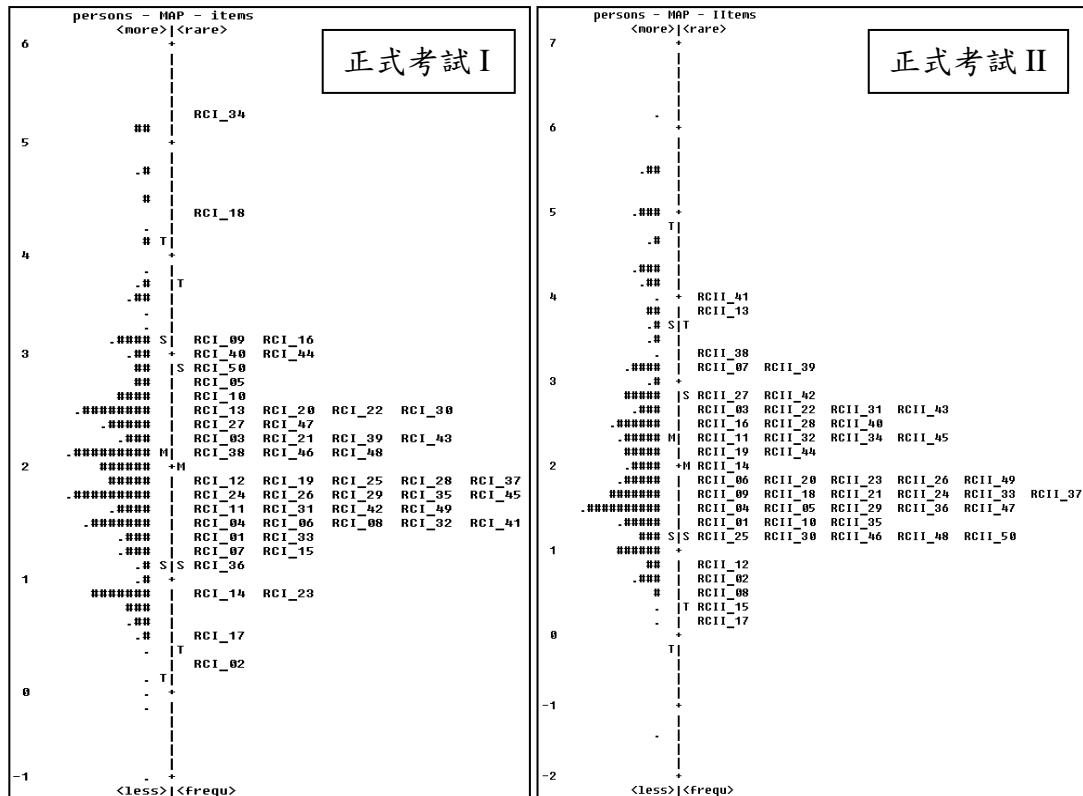


圖 10 流利精通級測驗受測者與試題分布圖

註：“#”在正式考試 I 與 II 分別表示三位及三位受測者；“.”表示一位受測者；RCI\_01~RCI\_50 和 RCII\_01~RCII\_50 表示試題編號。

由於預試時考生答題動機可能不若正式考試強烈，加上正式考試考生整體能力未必與試題難度相當，如入門基礎級考生整體能力即高於試題難度，適配度容易有偏低的現象，因此在正式考試採取較為寬鬆的標準，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 9 所示。入門基礎級兩次測驗各 50 道試題中，分別有六道及五道試題；進階高階級兩次測驗各 50 道試題中，分別有十道及六道試題；流利精通級兩次測驗各 50 道試題中，分別有兩道及一道試題。各等級測驗在兩次正式測驗的試題適配率介於 80% 至 98%，顯示適配比例大致良好。因為判斷試題不適配時，犯第一型錯誤（Type I error）的可能性為 5%，也就是說，若有 100 道試題進行適配度分析，仍有 5 題真正適配，但被判斷為不適配的可能性，流利精通級測驗的不適配題數也約略與此比例相當。

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 9 試題適配分布

測驗等級	入門基礎級		進階高階級		流利精通級	
	I	II	I	II	I	II
正式考試						
總試題數	50	50	50	50	50	50
適配題數	44	45	40	44	48	49
不適配題數	6	5	10	6	2	1
試題適配率(%)	88	90	80	88	96	98

進一步針對各等級測驗不適配試題進行選項分析。表 10、表 12 與表 14 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 10 可知，在入門基礎級兩次正式考試共 11 道不適配試題中，除 RAI\_39、RAII\_36、RAII\_37 與 RAII\_40 外，其餘七題選答正確選項的人數都是最多的，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。此外，試題中正確選項的點二系列值均符合標準<sup>4</sup>，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

<sup>4</sup> 點二系列相關係數標準為在.30 以上，惟若屬極端值難度試題，則標準放寬至.15 以上。入門基礎級試題難度參數低於-3.0 或高於-0.5、進階高階級難度參數低於-1.0 或高於 1.5、流利精通級難度參數低於 0.5 或高於 3.0，則標準為.15 以上。



效區別受測者能力之高低。

表 10 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	RAI_03	-3.683	0.202	A	0	9	-3.17	0.20	-.39
				C	0	10	-1.24	0.51	-.10
				B	1	414	-0.30	0.06	.35
	RAI_12	-2.787	0.154	C	0	17	-2.34	0.29	-.36
				B	0	84	-1.29	0.11	-.27
				A	1	332	-0.05	0.07	.42
	RAI_27	-2.798	0.155	A	0	13	-1.70	0.33	-.18
				B	0	79	-1.53	0.13	-.40
				C	1	341	-0.06	0.07	.45
	RAI_28	-2.713	0.151	C	0	13	-2.56	0.20	-.34
				A	0	21	-2.02	0.24	-.30
				B	1	399	-0.22	0.06	.46
	RAI_37	-2.429	0.142	C	0	37	-1.69	0.20	-.32
				B	0	92	-1.35	0.11	-.33
A				1	304	0.07	0.07	.49	
RAI_39	-0.535	0.112	B	0	43	-1.47	0.18	-.23	
			C	0	228	-0.45	0.08	-.30	
			A	1	162	0.01	0.12	.38	
RAII_06	-3.363	0.194	C	0	22	-2.24	0.29	-.30	
			A	0	47	-0.65	0.23	-.01	
			B	1	258	-0.24	0.10	.19	
			M	0	2	-1.61	0.66	-.04	
			C	0	22	-1.95	0.30	-.26	
			A	0	68	-1.64	0.17	-.35	
RAII_15	-2.737	0.171	B	1	237	0.07	0.10	.46	
			M	0	2	-3.30	1.02	-.15	
			B	0	87	-1.14	0.14	-.19	
			A	0	95	-0.83	0.17	-.12	
RAII_36	-0.596	0.138	C	1	146	0.26	0.13	.28	
			M	0	1	-4.32	--	-.16	
			B	0	38	-1.46	0.30	-.25	
RAII_37	-0.746	0.139	C	0	112	-0.91	0.13	-.12	
			A	1	176	0.13	0.12	.28	
			M	0	3	-2.82	0.76	-.15	
			B	0	44	-1.88	0.17	-.33	
RAII_40	-1.722	0.149	C	0	66	-0.81	0.18	-.05	
			A	1	217	0.00	0.11	.29	
			M	0	2	-4.10	0.22	-.21	

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。

檢視上述四題相關數值，RAI\_39 雖正答人數非最多，但因此題較難(-0.535)，且正答平均能力值為最高，且點二系列相關符合標準，故不進一步討論。研發人員檢視其餘三題的預試適配度分析結果，如表 11 所示，RAII\_36、RAII\_37 與 RAII\_40 預試後的適配結果皆符合模式。研發人員將持續觀察這些試題於正式考試後的分析結果，若未來於正式卷中仍持續發生不適配的情形，則考慮停止使用。

**表 11 入門基礎級不適配試題之預試適配分析結果**

	正式考試		預試	
	Infit MNSQ	Infit ZSTD	Infit MNSQ	Infit ZSTD
RAII_36	1.49	6.37	1.06	0.93
RAII_37	1.48	6.19	1.28	4.09
RAII_40	1.53	5.79	1.00	0.00

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，除了 RBI\_44、RBII\_12 與 RBII\_49 外，其餘 13 道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。此外，正答選項的點二系列相關皆符合標準，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。

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	RBI_06	-1.207	0.092	A	0	18	-0.64	0.21	-.17
				C	0	27	-0.38	0.16	-.16
				B	0	234	0.36	0.06	-.05
				D	1	585	0.66	0.04	.16
				M	0	2	-3.20	2.57	-.12
	RBI_10	-2.793	0.159	A	0	21	-0.26	0.19	-.11
				D	0	27	-0.24	0.20	-.12
				B	0	27	0.20	0.18	-.04
				C	1	787	0.58	0.04	.16
				M	0	4	-1.67	1.39	-.10
	RBI_11	-1.720	0.107	C	0	22	-0.47	0.16	-.16
				D	0	33	0.05	0.13	-.08
				B	0	145	0.07	0.07	-.16
				A	1	663	0.67	0.04	.24
				M	0	3	-1.32	2.24	-.05
	RBI_20	-1.168	0.091	C	0	142	-0.05	0.06	-.23
				D	0	76	0.02	0.08	-.14
				B	0	76	0.09	0.08	-.11
				A	1	566	0.78	0.04	.33
				M	0	6	-0.84	1.00	-.07
	RBI_35	-1.291	0.094	C	0	15	-0.88	0.20	-.18
				A	0	10	-0.80	0.16	-.14
				B	0	19	-0.37	0.22	-.13
				D	1	818	0.58	0.03	.27
				M	0	4	-1.51	1.54	-.09
	RBI_36	-0.938	0.086	B	0	17	-0.85	0.19	-.19
				A	0	61	-0.54	0.08	-.29
				D	0	17	-0.38	0.15	-.12
				C	1	767	0.65	0.03	.38
				M	0	4	-1.51	1.54	-.09
	RBI_44	2.125	0.092	C	0	176	0.38	0.06	-.05
				A	0	184	0.43	0.07	-.02
				D	0	211	0.45	0.05	.00
				B	1	275	0.74	0.08	.06
				M	0	20	-0.23	0.33	-.09
	RBI_45	1.843	0.087	C	0	120	0.16	0.07	-.13
				B	0	166	0.25	0.07	-.11
				A	0	212	0.37	0.06	-.05
				D	1	351	0.88	0.06	.23
				M	0	17	-0.39	0.37	-.10
	RBI_47	1.777	0.085	A	0	142	0.18	0.07	-.13
				B	0	189	0.27	0.05	-.11
				D	0	178	0.34	0.06	-.06
				C	1	335	0.92	0.06	.25
				M	0	22	-0.24	0.31	-.10
	RBI_50	2.187	0.094	B	0	116	0.09	0.07	-.17
				A	0	157	0.32	0.06	-.07
C				0	243	0.48	0.05	.01	
D				1	296	0.85	0.07	.17	
M				0	54	0.18	0.15	-.06	

表 12(續) 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
II	RBII_12	-0.267	0.090	A	0	48	0.05	0.09	-0.20
				B	0	75	0.53	0.13	-0.10
				D	0	164	0.79	0.08	-0.01
				C	1	432	1.10	0.05	.18
				M	0	8	-0.36	0.47	-0.12
	RBII_18	-2.158	0.151	C	0	4	-0.78	0.57	-0.11
				A	0	9	-0.41	0.21	-0.13
				D	0	7	-0.39	0.24	-0.11
				B	1	707	0.93	0.04	.21
	RBII_19	-1.529	0.121	C	0	8	-0.46	0.43	-0.13
				B	0	19	-0.21	0.20	-0.16
				D	0	13	-0.20	0.24	-0.13
				A	1	686	0.96	0.04	.24
				M	0	1	-0.14	--	-0.03
	RBII_29	-0.656	0.096	C	0	80	0.09	0.10	-0.25
				D	0	51	0.13	0.12	-0.18
				A	0	162	0.45	0.07	-0.18
				B	1	431	1.30	0.05	.41
				M	0	3	-0.08	0.97	-0.06
	RBII_47	2.258	0.095	A	0	94	0.41	0.09	-0.17
B				0	130	0.59	0.08	-0.12	
D				0	185	0.95	0.07	.06	
C				1	290	1.23	0.07	.16	
M				0	28	0.00	0.17	-0.16	
RBII_49	2.191	0.094	A	0	130	0.76	0.08	-0.05	
			C	0	199	0.82	0.07	-0.04	
			D	0	124	1.05	0.10	.07	
			B	1	240	1.04*	0.07	.03	
			M	0	34	0.17	0.17	-0.14	

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；“\*”表示選答誘答選項考生平均能力值高於選正確選項考生。

研發人員檢視上述三題的預試適配度分析結果，如表 13 所示，RBI\_44、RBII\_12 與 RBII\_49 預試後的適配結果皆符合模式。研發人員將持續觀察這些試題於正式考試後的分析結果，若未來於正式卷中仍持續發生不適配的情形，則考慮停止使用。

表 13 進階高階級不適配試題之預試適配分析結果

	正式考試		預試	
	Infit MNSQ	Infit ZSTD	Infit MNSQ	Infit ZSTD
RBI_44	1.67	9.90	1.10	1.82
RBII_12	1.44	9.55	1.13	2.39
RBII_49	1.59	9.90	1.17	2.86

表 14 為流利精通級測驗選項分析結果，三道不適配試題，除 RCI\_18 外，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相關數值符合標準，此即表示測驗具有幅合效度；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。

RCI\_18 此次模式適配結果，Infit MNSQ 與 Infit ZSTD 分別為 1.79 與 5.35，檢視預試時的分析結果，兩項指標數值分別為 1.32 與 2.55，MNSQ 略微超出標準，經研發人員就試題內容討論後決議保留。研發人員將持續觀察此題於正式考試後的分析結果，若未來於正式卷中仍持續發生不適配的情形，則考慮停止使用。

表 14 流利精通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	RCI_02	0.225	0.152	D	0	24	1.27	0.12	-.23
				A	0	34	1.52	0.12	-.18
				C	0	33	2.16	0.17	.05
				B	1	244	2.25	0.06	.22
				M	0	3	-1.09	1.15	-.20
	RCI_18	4.358	0.177	C	0	65	1.82	0.11	-.12
				B	0	89	1.98	0.08	-.04
				D	0	115	2.24	0.10	.13
				A	1	66	2.23*	0.15	.01
				M	0	3	-0.01	1.68	-.11
II	RCII_50	1.214	0.137	D	0	20	1.46	0.16	-.21
				A	0	33	0.88	0.14	-.13
				B	0	77	1.90	0.09	-.21
				C	1	155	2.90	0.11	.37
				M	0	20	1.54	0.22	-.15

註：M 表示沒有填答；“\*”表示選答誘答選項考生平均能力值高於選正確選項考生。

## (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(Muthén & Muthén, 2012)，來檢視閱讀測驗入門基礎級、進階高階級與流利精通級雙向細目表中欲測得之閱讀分項能力，是否可經由受測者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版進行類別變項之驗證性因素分析(CFA with categorical factor indicators)，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)，各測驗等級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(1988)以及Hu和Bentler(1998)的研究，研究者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( $\chi^2/df$ )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、進階高階級與流利精通級5月和11月正式考試驗證性因素分析結果如下圖11至圖13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1至3。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.15至.82之間，平均值為.63；因素負荷量標準誤為0.04至0.10；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.91、.90及.93。11月正式考試因素負荷量介於.30至.93之間，平均值為.70；因素負荷量標準誤為0.02至0.09；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.97、.97及1.01。

進階高階級測驗四因素與五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.09至.81之間，平均值為.45；因素負荷量標準誤為0.03至0.07；除了RBI\_44外，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.76至.92之間。11月正式考試因素負荷量介於.06至.79之間，平均值為.52；因素負荷量標準誤為0.03至0.12；除了RBII\_49外，因素負荷量統計考驗均達顯著水準( $p < .05$ )；五個潛在變項間的相關係數介於.55至.95之間。

流利精通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.02至.67之間，平均值為.48；因素負荷量標準誤為0.05至0.11；除了RCI\_18、RCI\_29與RCI\_34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.91。11月正式考試因素負荷量介於.03至.92之間，平均值為.56；因素負荷量標準誤為0.04至0.10；除了RCII\_15與RCII\_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。

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。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的試題比例分別為2%與0%，進階高階級和流利精通級依序為14%、6%、10%及12%，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。

測驗研發人員與統計分析人員針對分析結果中因素負荷量統計考驗未達.05顯著水準的試題進行討論，在進階高階級部分，討論結果認為，可能與實際所測得之理解能力層次較高有關，RBI\_44 與 RBII\_49 測得了根據閱讀文本進行抽象推論的能力，與進階高階級其他主要在測知能否尋找較為直接、明確的文本訊息或細節之能力的試題較為不同。至於流利精通級方面，兩次正式考試因素負荷量統計考驗不顯著的試題，經研發人員歸納發現，RCI\_29 與 RCI\_34 可能與文本主題較為抽象且作答需要推論有關。而 RCII\_41 除了文本主題抽象外，文本長度也較其他題組略長，該題的提問設計角度與其他試題較為不同，更多地涉及了考生在閱讀文本時立場的轉換，方便因素負荷量較低。然而另一方面，在流利精通級測驗的能力描述中，實包含了「讀懂言外之意」此一推論能力，又，其他同樣在測知推論能力的試題其因素負荷量考驗數值仍達到顯著水準，這或許顯示了推論能力可以再細分為不同層次或程度，因此未來研發人員將持續對因素負荷量較低的試題內容進行觀察與討論。

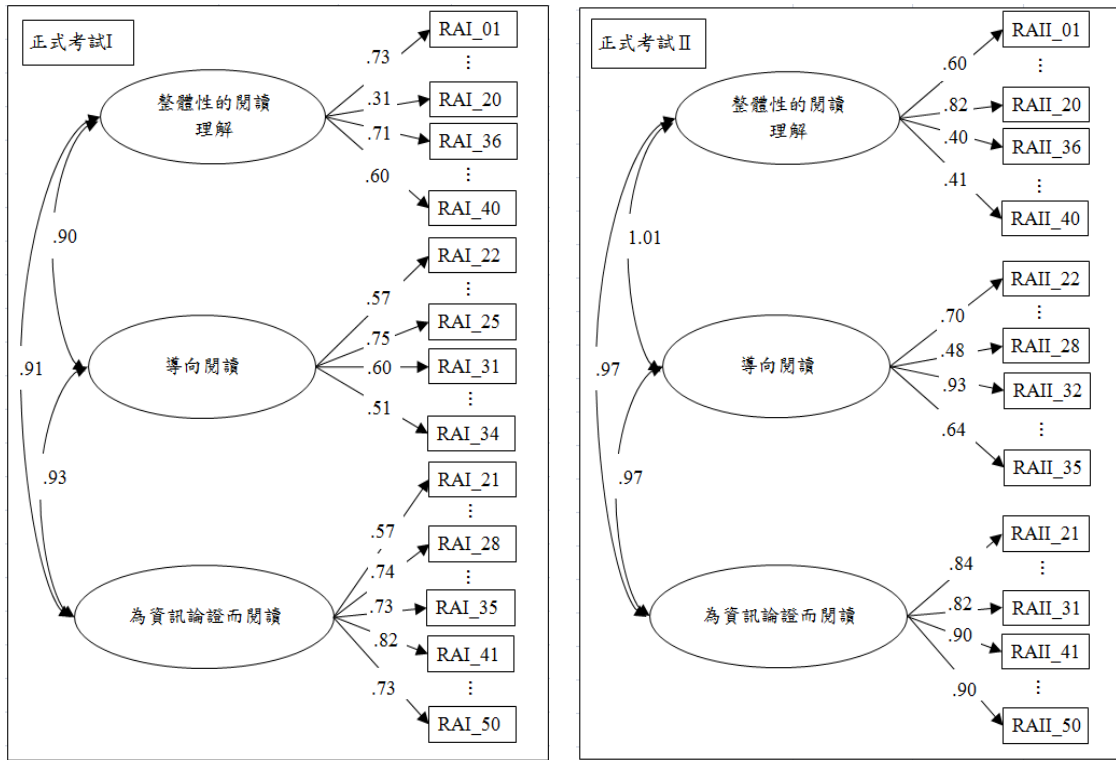


圖 11 入門基礎級閱讀測驗驗證性因素分析

註：RAI\_01-RAI\_20、RAI\_36-RAI\_40、RAII\_01-RAII\_20、RAII\_36-RAII\_40所屬分項能力為「整體性的閱讀理解」；RAI\_22-RAI\_23、RAI\_25、RAI\_31-RAI\_34、RAII\_22-RAII\_28、RAII\_32-RAII\_35所屬分項能力為「導向閱讀」；RAI\_21、RAI\_24、RAI\_26-RAI\_30、RAI\_35、RAI\_41-RAI\_50、RAII\_21、RAII\_29-RAII\_31、RAII\_41-RAII\_50所屬分項能力為「為資訊論證而閱讀」。



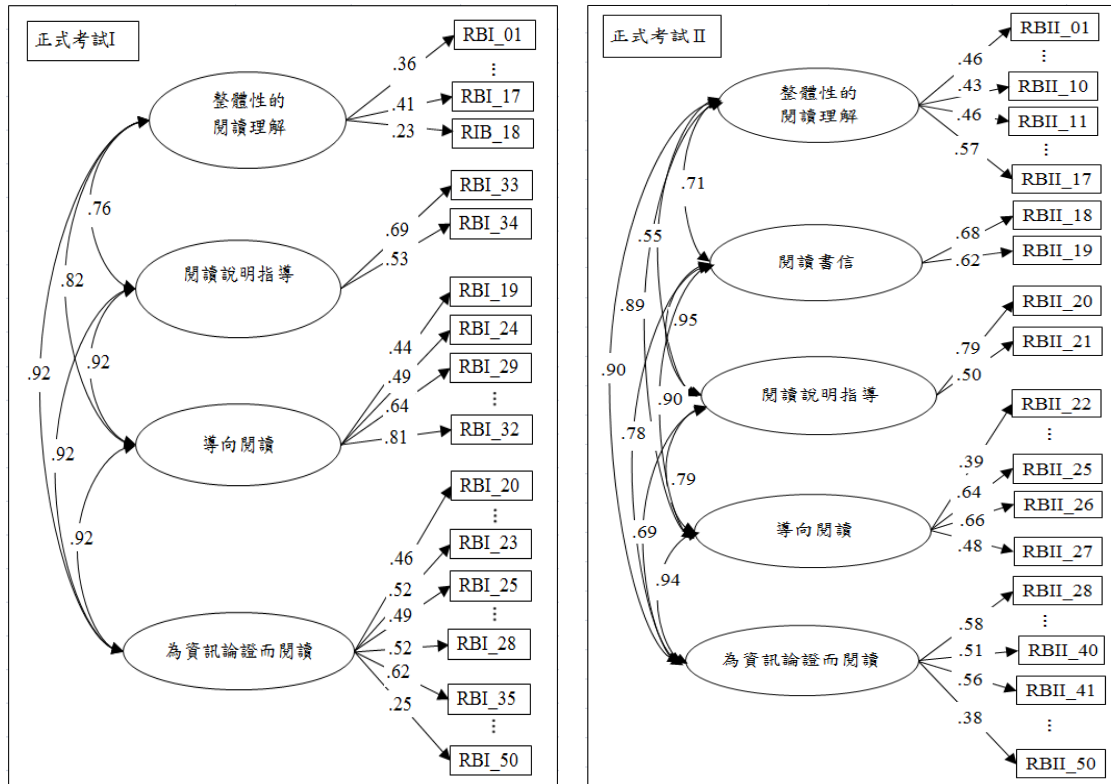


圖 12 進階高階級閱讀測驗驗證性因素分析

註：RBI\_01-RBI\_18、RBII\_01-RBII\_17所屬分項能力為「整體性的閱讀理解」；RBI\_19、RBI\_24、RBI\_29-RBI\_32、RBII\_22-RBII\_27所屬分項能力為「導向閱讀」；RBI\_33-RBI\_34、RBII\_20-RBII\_21所屬分項能力為「閱讀說明指導」；RBII\_18-RBII\_19所屬分項能力為「閱讀書信」；RBI\_20-RBI\_23、RBI\_25-RBI\_28、RBI\_35-RBI\_50、RBII\_28-RBII\_50所屬分項能力為「為資訊論證而閱讀」。

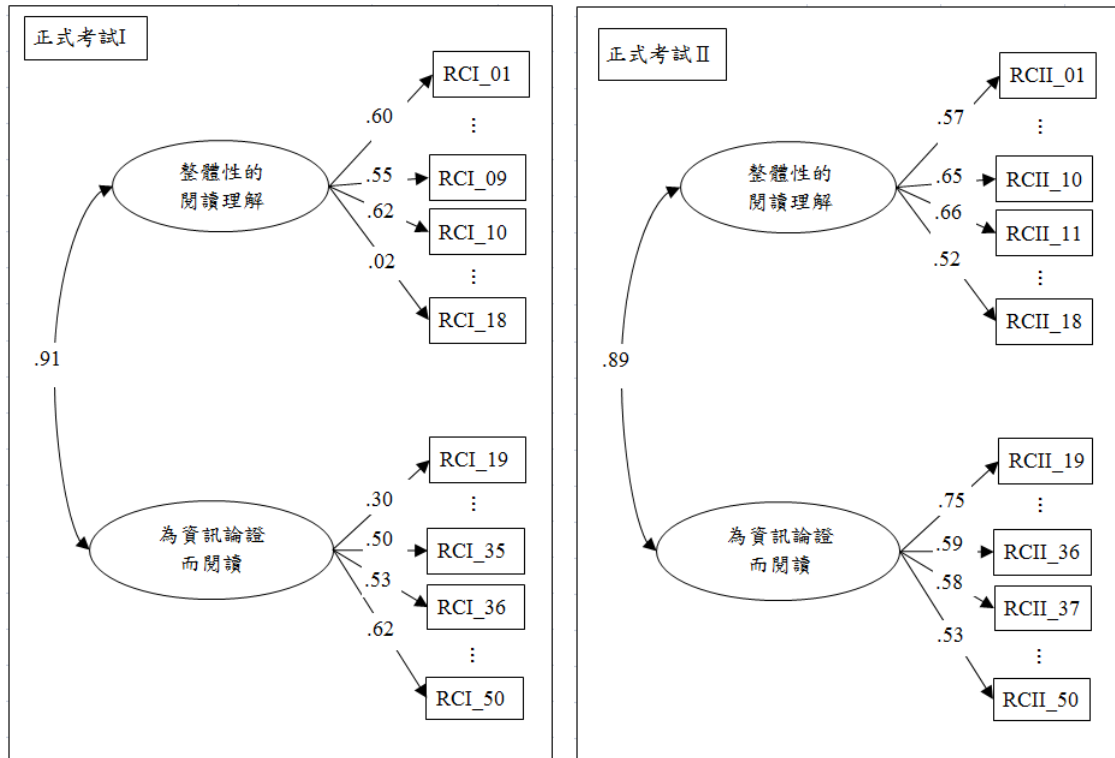


圖 13 流利精通級閱讀測驗驗證性因素分析

註：RCI\_01-RCI\_18、RCII\_01-RCII\_18 所屬分項能力為「整體性的閱讀理解」；RCI\_19-RCI\_50、RCII\_19-RCII\_50 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( $\chi^2/df$ )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 15 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 3，表示模式適配度良好(Wheaton, Muthen, Alwin & Summers, 1977)。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，除了進階高階級 5 月正式考試的 NNFI 為 .89 外，CFI 和 NNFI 數值皆在 .90 以上，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，各測驗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 15 整體模式適配度指標摘要及模式比較表**

測驗等級	正式考試	模式	$\chi^2$	$\chi^2/df$	RMSEA	CFI	NNFI
入門基礎級	I	三因素模式	1382.05**	1.18	.02	.96	.96
	II	三因素模式	1285.31*	1.10	.02	.99	.99
進階高階級	I	四因素模式	1939.81**	1.66	.03	.90	.89
	II	五因素模式	1553.43**	1.33	.02	.96	.96
流利精通級	I	二因素模式	1298.99**	1.11	.02	.97	.96
	II	二因素模式	1319.47**	1.12	.02	.98	.97

\* $p < .05$  ; \*\* $p < .01$

## 五、 結論

本文為 2014 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、通過門檻等方面進行概述，並說明測驗研發、施測及成績公布之標準化流程。第二部分的內容則以評估 2014 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。同樣測量華語文能力的新漢語水平考試 (HSK)，在 2011 年公布的分析結果顯示，一級聽力與閱讀測驗合計的信度係數介於 0.85-0.95 之間，三級聽讀合計的信度在 0.90-0.95 之間；五級聽讀合計的信度在 0.90-0.95 之間(張晉軍，2011)。其他大型語言測驗，如 TOEIC 於 2012 年的測驗介紹手冊宣稱聽讀測驗 KR20 信度係數接近或達到 0.90 以上 (ETS，2012)。顯見閱讀測驗信度分析結果與相關語言測驗的數值相仿。

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中使用，故專家審查程序極為重要。為提供更為嚴謹、全面的專家效度證據，本會於本年度設計了「試題預設等級評估量表」，請專家學者針對每一道試題詳細評估，希望透過提供外審專家更具針對性的試題檢核材料，來記錄並量化專家的試題評估結果。與此同時，因考量到命題教師培訓也是保障、提升試題品質不可忽視的一環，本測驗未來也將在命題教師執行命題任務的流程中，增列「命題原則檢核及自評」一項步驟，以便審題人員逐步調整並指派後續之命題任務，進而獲得命題培訓方面的效度證據。

本測驗除了具備測驗之內容效度證據之外，在施測完成後，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是否相同，並以此作為本測驗之建構效度證據。本年度另進行分項能力與測驗題型兩種測量

架構之比較，未來將根據分析結果，採用測驗題型作為驗證性因素分析的測量架構。

綜言之，根據本文之信度、效度分析結果顯示，本會於 2014 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。
- 國家華語測驗推動工作委員會(2015)。華語文能力測驗技術報告 2013 (2) 閱讀測驗信效度(ISBN 978-986-92167-3-9)。新北市：國家華語測驗推動工作委員會。
- 張晉軍(2011)。新漢語水準考試(HSK)品質報告。2013 年 9 月 26 日，取自：  
[http://blog.sina.com.cn/s/blog\\_53e7c11d0100v7lz.html](http://blog.sina.com.cn/s/blog_53e7c11d0100v7lz.html)
- 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。
- 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。
- 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.
- 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](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.
- Educational Testing Service. (2012). *TOEIC Examinee handbook listening & reading*. Retrieved Sep 26, 2013, from [http://www.ets.org/Media/Tests/TOEIC/pdf/TOEIC\\_LR\\_examinee\\_handbook.pdf](http://www.ets.org/Media/Tests/TOEIC/pdf/TOEIC_LR_examinee_handbook.pdf)
- 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.
- 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. (2012). Mplus (Version 7.0) [Computer Software]. 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.)
- Wheaton, B., Muthen, B., Alwin, D., & Summers, G. (1977). Assessing the reliability and stability in panel models. In D.R. Heise (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.

附件 1 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	RAI_01	.73	0.07	.00
		RAI_02	.57	0.10	.00
		RAI_03	.70	0.10	.00
		RAI_04	.62	0.09	.00
		RAI_05	.64	0.08	.00
		RAI_06	.60	0.08	.00
		RAI_07	.74	0.05	.00
		RAI_08	.70	0.07	.00
		RAI_09	.73	0.06	.00
		RAI_10	.70	0.06	.00
		RAI_11	.57	0.07	.00
		RAI_12	.60	0.06	.00
		RAI_13	.54	0.07	.00
		RAI_14	.57	0.07	.00
		RAI_15	.68	0.06	.00
		RAI_16	.71	0.05	.00
		RAI_17	.77	0.05	.00
		RAI_18	.51	0.06	.00
		RAI_19	.49	0.06	.00
		RAI_20	.31	0.07	.00
	RAI_36	.71	0.06	.00	
	RAI_37	.67	0.05	.00	
	RAI_38	.67	0.07	.00	
	RAI_39	.15	0.07	.04	
	RAI_40	.60	0.08	.00	
	RAI_22	.57	0.09	.00	
	RAI_23	.68	0.05	.00	
	RAI_25	.75	0.07	.00	
	RAI_31	.60	0.05	.00	
	RAI_32	.61	0.06	.00	
	RAI_33	.74	0.05	.00	
	RAI_34	.51	0.05	.00	
	RAI_21	.57	0.05	.00	
	RAI_24	.76	0.05	.00	
	RAI_26	.68	0.04	.00	
	RAI_27	.65	0.06	.00	
	RAI_28	.74	0.06	.00	
	RAI_29	.61	0.06	.00	
	RAI_30	.66	0.06	.00	
	RAI_35	.73	0.06	.00	
	RAI_41	.82	0.05	.00	
	RAI_42	.70	0.05	.00	
	RAI_43	.50	0.05	.00	
	RAI_44	.58	0.07	.00	
	RAI_45	.50	0.07	.00	
	RAI_46	.66	0.06	.00	
	RAI_47	.70	0.05	.00	
	RAI_48	.73	0.06	.00	
	RAI_49	.67	0.05	.00	
	RAI_50	.73	0.06	.00	



正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
II	整體性的閱讀理解	RAII_01	.60	0.08	.00
		RAII_02	.58	0.09	.00
		RAII_03	.75	0.05	.00
		RAII_04	.55	0.08	.00
		RAII_05	.75	0.05	.00
		RAII_06	.30	0.08	.00
		RAII_07	.59	0.07	.00
		RAII_08	.67	0.07	.00
		RAII_09	.72	0.05	.00
		RAII_10	.70	0.06	.00
		RAII_11	.72	0.05	.00
		RAII_12	.74	0.05	.00
		RAII_13	.83	0.04	.00
		RAII_14	.74	0.05	.00
		RAII_15	.63	0.06	.00
		RAII_16	.68	0.05	.00
		RAII_17	.74	0.05	.00
		RAII_18	.73	0.05	.00
		RAII_19	.73	0.05	.00
		RAII_20	.82	0.04	.00
	RAII_36	.40	0.07	.00	
	RAII_37	.39	0.07	.00	
	RAII_38	.69	0.05	.00	
	RAII_39	.47	0.07	.00	
	RAII_40	.41	0.07	.00	
	導向閱讀	RAII_22	.70	0.06	.00
		RAII_23	.86	0.03	.00
		RAII_24	.80	0.04	.00
		RAII_25	.63	0.06	.00
		RAII_26	.73	0.05	.00
		RAII_27	.79	0.05	.00
		RAII_28	.48	0.07	.00
		RAII_32	.93	0.02	.00
		RAII_33	.72	0.05	.00
		RAII_34	.71	0.05	.00
	RAII_35	.64	0.06	.00	
	為資訊論證而閱讀	RAII_21	.84	0.04	.00
		RAII_29	.81	0.05	.00
		RAII_30	.78	0.05	.00
		RAII_31	.82	0.04	.00
		RAII_41	.90	0.03	.00
		RAII_42	.64	0.05	.00
		RAII_43	.69	0.05	.00
		RAII_44	.78	0.05	.00
		RAII_45	.72	0.05	.00
		RAII_46	.77	0.04	.00
	RAII_47	.87	0.03	.00	
	RAII_48	.70	0.05	.00	
	RAII_49	.85	0.04	.00	
	RAII_50	.90	0.03	.00	

附件 2 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	RBI_01	.36	0.05	.00
		RBI_02	.41	0.05	.00
		RBI_03	.36	0.05	.00
		RBI_04	.66	0.04	.00
		RBI_05	.42	0.06	.00
		RBI_06	.24	0.05	.00
		RBI_07	.36	0.04	.00
		RBI_08	.51	0.04	.00
		RBI_09	.49	0.05	.00
		RBI_10	.34	0.07	.00
		RBI_11	.37	0.05	.00
		RBI_12	.41	0.04	.00
		RBI_13	.46	0.05	.00
		RBI_14	.55	0.05	.00
		RBI_15	.43	0.05	.00
		RBI_16	.52	0.05	.00
		RBI_17	.41	0.04	.00
		RBI_18	.23	0.05	.00
	導向閱讀	RBI_19	.44	0.05	.00
		RBI_24	.49	0.04	.00
		RBI_29	.64	0.04	.00
		RBI_30	.57	0.04	.00
		RBI_31	.74	0.03	.00
		RBI_32	.81	0.03	.00
	閱讀說明指導	RBI_33	.69	0.06	.00
		RBI_34	.53	0.05	.00
	為資訊論證而閱讀	RBI_20	.46	0.04	.00
		RBI_21	.43	0.05	.00
		RBI_22	.65	0.04	.00
		RBI_23	.52	0.04	.00
		RBI_25	.49	0.04	.00
		RBI_26	.46	0.04	.00
		RBI_27	.23	0.04	.00
		RBI_28	.52	0.04	.00
		RBI_35	.62	0.06	.00
		RBI_36	.68	0.04	.00
		RBI_37	.56	0.05	.00
		RBI_38	.40	0.04	.00
		RBI_39	.63	0.04	.00
		RBI_40	.22	0.05	.00
		RBI_41	.40	0.05	.00
		RBI_42	.28	0.04	.00
		RBI_43	.40	0.05	.00
		RBI_44	.09	0.05	.07
	RBI_45	.33	0.04	.00	
	RBI_46	.54	0.04	.00	
	RBI_47	.35	0.04	.00	
	RBI_48	.32	0.04	.00	
	RBI_49	.30	0.05	.00	
	RBI_50	.25	0.05	.00	

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
II	整體性的閱讀理解	RBII_01	.46	0.04	.00
		RBII_02	.51	0.04	.00
		RBII_03	.46	0.05	.00
		RBII_04	.37	0.05	.00
		RBII_05	.67	0.04	.00
		RBII_06	.42	0.04	.00
		RBII_07	.34	0.06	.00
		RBII_08	.79	0.03	.00
		RBII_09	.34	0.05	.00
		RBII_10	.43	0.05	.00
		RBII_11	.46	0.04	.00
		RBII_12	.27	0.05	.00
		RBII_13	.69	0.04	.00
		RBII_14	.61	0.04	.00
		RBII_15	.44	0.05	.00
		RBII_16	.50	0.04	.00
		RBII_17	.57	0.04	.00
	閱讀書信	RBII_18	.68	0.12	.00
		RBII_19	.62	0.11	.00
	閱讀說明指導	RBII_20	.79	0.09	.00
		RBII_21	.50	0.08	.00
	導向閱讀	RBII_22	.39	0.05	.00
		RBII_23	.40	0.05	.00
		RBII_24	.47	0.05	.00
		RBII_25	.64	0.04	.00
		RBII_26	.66	0.04	.00
		RBII_27	.48	0.05	.00
	為資訊論證而閱讀	RBII_28	.58	0.04	.00
		RBII_29	.56	0.04	.00
		RBII_30	.42	0.05	.00
		RBII_31	.64	0.04	.00
		RBII_32	.46	0.04	.00
		RBII_33	.49	0.05	.00
		RBII_34	.59	0.04	.00
		RBII_35	.74	0.03	.00
		RBII_36	.61	0.04	.00
		RBII_37	.57	0.04	.00
		RBII_38	.43	0.04	.00
		RBII_39	.69	0.03	.00
		RBII_40	.51	0.04	.00
		RBII_41	.56	0.04	.00
		RBII_42	.67	0.04	.00
		RBII_43	.51	0.04	.00
		RBII_44	.78	0.03	.00
		RBII_45	.58	0.04	.00
		RBII_46	.45	0.05	.00
		RBII_47	.26	0.05	.00
		RBII_48	.53	0.05	.00
		RBII_49	.06	0.05	.22
		RBII_50	.38	0.05	.00

附件 3 流利精通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	RCI_01	.60	0.05	.00
		RCI_02	.37	0.07	.00
		RCI_03	.57	0.06	.00
		RCI_04	.57	0.05	.00
		RCI_05	.52	0.06	.00
		RCI_06	.51	0.06	.00
		RCI_07	.48	0.06	.00
		RCI_08	.60	0.06	.00
		RCI_09	.55	0.07	.00
		RCI_10	.62	0.06	.00
		RCI_11	.39	0.06	.00
		RCI_12	.50	0.06	.00
		RCI_13	.48	0.07	.00
		RCI_14	.49	0.06	.00
		RCI_15	.50	0.06	.00
		RCI_16	.30	0.08	.00
		RCI_17	.67	0.05	.00
		RCI_18	.02	0.09	.86
	為資訊論證而閱讀	RCI_19	.30	0.07	.00
		RCI_20	.42	0.06	.00
		RCI_21	.49	0.06	.00
		RCI_22	.58	0.05	.00
		RCI_23	.44	0.07	.00
		RCI_24	.51	0.06	.00
		RCI_25	.42	0.06	.00
		RCI_26	.56	0.05	.00
		RCI_27	.55	0.06	.00
		RCI_28	.30	0.07	.00
		RCI_29	.13	0.07	.07
		RCI_30	.25	0.07	.00
		RCI_31	.55	0.05	.00
		RCI_32	.65	0.05	.00
		RCI_33	.61	0.05	.00
		RCI_34	.09	0.11	.39
		RCI_35	.50	0.06	.00
		RCI_36	.53	0.06	.00
		RCI_37	.58	0.05	.00
		RCI_38	.51	0.06	.00
		RCI_39	.51	0.06	.00
		RCI_40	.21	0.08	.01
		RCI_41	.51	0.06	.00
		RCI_42	.57	0.05	.00
		RCI_43	.55	0.05	.00
		RCI_44	.38	0.07	.00
		RCI_45	.67	0.05	.00
		RCI_46	.50	0.06	.00
		RCI_47	.60	0.05	.00
		RCI_48	.54	0.06	.00
		RCI_49	.49	0.06	.00
		RCI_50	.62	0.05	.00

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
II	整體性的閱讀理解	RCII_01	.57	0.05	.00
		RCII_02	.62	0.05	.00
		RCII_03	.69	0.05	.00
		RCII_04	.55	0.05	.00
		RCII_05	.50	0.06	.00
		RCII_06	.66	0.05	.00
		RCII_07	.92	0.04	.00
		RCII_08	.62	0.05	.00
		RCII_09	.64	0.05	.00
		RCII_10	.65	0.04	.00
		RCII_11	.66	0.05	.00
		RCII_12	.62	0.05	.00
		RCII_13	.83	0.06	.00
		RCII_14	.58	0.05	.00
		RCII_15	.15	0.08	.08
		RCII_16	.50	0.06	.00
		RCII_17	.68	0.04	.00
		RCII_18	.52	0.06	.00
	為資訊論證而閱讀	RCII_19	.75	0.04	.00
		RCII_20	.56	0.05	.00
		RCII_21	.52	0.06	.00
		RCII_22	.29	0.07	.00
		RCII_23	.60	0.05	.00
		RCII_24	.70	0.04	.00
		RCII_25	.56	0.05	.00
		RCII_26	.80	0.04	.00
		RCII_27	.69	0.05	.00
		RCII_28	.51	0.06	.00
		RCII_29	.57	0.05	.00
		RCII_30	.56	0.05	.00
		RCII_31	.70	0.05	.00
		RCII_32	.60	0.06	.00
		RCII_33	.47	0.06	.00
		RCII_34	.56	0.06	.00
		RCII_35	.43	0.06	.00
		RCII_36	.59	0.05	.00
		RCII_37	.58	0.06	.00
		RCII_38	.24	0.08	.00
		RCII_39	.55	0.08	.00
		RCII_40	.51	0.06	.00
		RCII_41	.03	0.10	.76
		RCII_42	.27	0.07	.00
		RCII_43	.28	0.07	.00
		RCII_44	.48	0.06	.00
		RCII_45	.67	0.05	.00
		RCII_46	.57	0.05	.00
		RCII_47	.73	0.04	.00
		RCII_48	.71	0.05	.00
		RCII_49	.54	0.06	.00
		RCII_50	.53	0.06	.00



書名：華語文能力測驗技術報告—2014(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：2016 年 11 月

定價：新台幣 100 元

版權所有

翻印必究

