

華語文能力測驗技術報告 — 2013(1)

聽力測驗信效度

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) 題庫輸入.....	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.....	32
六、	文獻.....	33

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	16
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	25
表 13	整體模式適配度指標摘要表	31

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	18
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分布圖	20
圖 11	LIIA_35 試題特徵曲線	24
圖 12	LIB_27 試題特徵曲線	25
圖 13	LIB_48 試題特徵曲線	26
圖 14	LIIB_50 試題特徵曲線	26
圖 15	入門基礎級測驗二因素驗證性因素分析	28
圖 16	進階高階級測驗三因素驗證性因素分析	29
圖 17	流利級測驗三因素驗證性因素分析	30

附件目錄

附件 1	聽力測驗各等級標準設定會議問卷調查結果	35
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	36
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	38
附件 4	流利級測驗各題因素負荷量及殘差變異量摘要表	40

一、前言

「華語文聽力測驗」(以下簡稱聽力測驗)為一套專為母語非華語者所研發的標準化語言能力測驗，從 2001 年 8 月開始籌劃、2003 年 12 月正式對外開辦考試。為使本測驗能夠與國際語言標準接軌，提供考生能夠衡量其語言能力的國際評量工具，本會於 2008 年起積極著手聽力測驗的改版工作¹。

自 2013 年起，華語文聽力測驗的測驗架構將語言能力分成三等六級，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。此測驗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 年華語文聽力測驗簡介、測驗實施、通過門檻制定程序與製卷、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度及效度分析結果。最後，根據各項分析結果提出相關討論及建議。

¹ 各等級聽力測驗分別對應於歐洲理事會(2001)所頒佈的「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 年華語文聽力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)；以及流利級(Level 5)。並分別對應於歐洲共同語文參考架構(CEFR) A1 (Breakthrough)、A2 (Waystage)、B1 (Threshold)、B2 (Vantage) 與 C1 (Effective Operational Proficiency)。以下就各等級能力描述、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聽力能力如表 1 所示。其中，入門基礎級著重「在日常生活的一般簡易溝通能力」；進階高階級著重在「語言段落的理解分析能力」；流利級則著重在「語言使用的廣度與精熟度」。

取得華語文聽力測驗各級證書，所代表之語言能力說明如下：

表 1 基本能力描述

通過等級	能力描述
入門級	在對方說話緩慢且清晰，並隨時準備提供協助的前提下，能聽懂基本短語或常用詞。
基礎級	在言語清楚且緩慢的情況下，對於簡短的言談，能理解內容與掌握重點。
進階級	當談話內容為與工作、學習、娛樂相關的熟悉話題，且講話人的口齒清晰、語音標準時，能了解內容大意和重要細節。
高階級	對於具有一定篇幅且以標準語表達的談話內容，包括專攻領域的技術性討論，不論內容抽象與否，都能聽懂要點大意。
流利級	能聽懂各種抽象或複雜主題的話語內容，即使話語結構或關聯性可能不夠清楚、明確；但在不熟悉說話人口音的情況下，可能需要特別確認部分細節。

(二) 測驗題型與題數

聽力理解能力並非只是被動地接收信息，而是涉及了一連串主動解碼推理的複雜認知過程。當聽者接收到一串語音訊息時，通過內化的聲音感應，把詞彙、

語法等語言形式的意義和語音迅速匹配、組合並處理資訊，再通過比較、聯想、預測來推斷接下來可能出現的訊息。聽者需要運用各種修正和補償策略(語言和非語言的知識)以達到理解，並完成溝通交際的任務。

入門基礎級聽力測驗為本測驗之起點測驗，此階段的語言使用者僅具有初階的語言能力，需倚賴大量的視覺輔助及語言之外的協助才能了解所接收到的訊息。因此，針對此階段受測者採用圖像式測驗，即使用大量圖片作為視覺輔助，透過圖片呈現任務。

進階高階級的語言使用者已能掌握日常生活中常見的，或是工作、學習以及娛樂等熟悉的話題，以及公共廣播、宣布等以標準語表達的內容，能了解這些訊息的大意和重要細節。而此階段程度較佳的學習者，對於具有一定篇幅的談話內容，不論內容是否抽象，都能聽懂大意與訊息要點。故在此等級題型設計中，除沿用入門基礎級問答理解(單輪)和對話理解(雙輪)的對話形式外，再加入較長篇幅的對話和段落形式，題幹及選項的呈現上皆以文字取代圖片設計。

對於流利級的語言使用者而言，其已經具備高度穩定的語言能力，能夠處理所有與日常生活相關的話題，不管話語結構或關聯性是否清楚或明確，都能聽懂並掌握要點與重要細節。因此針對此階段受測者，不論是對話或是段落形式的篇幅均加長，內容取材的範圍也更深入至專業議題。

各等級之題型與題數分布如表 2 所示，所有測驗試題皆為選擇題。其中，入門基礎級聽力測驗分為三個部分：看圖回答、問答理解、對話理解，共 50 道三選一的單選題，考試時間約為 60 分鐘。進階高階級和流利級則分為二個部分：對話與段落，共 50 道四選一的單選題，考試時間約為 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	看圖回答	50
	問答理解	
	對話理解	
進階高階級	對話	50
	段落	
流利級	對話	50
	段落	

(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)之相關研究。

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 道聽力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，在此為受測者能力值。

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，六個等級第三回合的標準差均低於前兩回合。表示經過討論後，判斷結果的一致性逐漸提高。從相關分析結果可發現，第三回合聽力測驗六個等級的相關係數介於.557 至.763 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中高度的相關。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	0.528	0.345	0.305
	基礎級	0.718	1.202	0.553
進階高階級	進階級	0.557	0.281	0.242
	高階級	0.622	0.289	0.254
流利精通級	流利級	1.104	0.293	0.234
	精通級	1.294	0.540	0.365

華語文聽力測驗入門基礎級、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，表示新版華語文聽力測驗，將華語學習者依其報考等級與測驗表現，區分為 CEFR A1 至 C2 六個等級的有效性得到驗證。

依據上述標準設定結果，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，以量尺分數(scale score)呈現。此做法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聽力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	通過等級	通過分數
入門基礎級	入門級	41
	基礎級	60
進階高階級	進階級	46
	高階級	61
流利級 ³	流利級	50

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

聽力測驗標準化流程共包含兩個部分(如圖 1 所示)，第一部分為考試前之正式考試製卷，分別為：試題收集、修審、預試、分析、題庫輸入、組合正式卷、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。

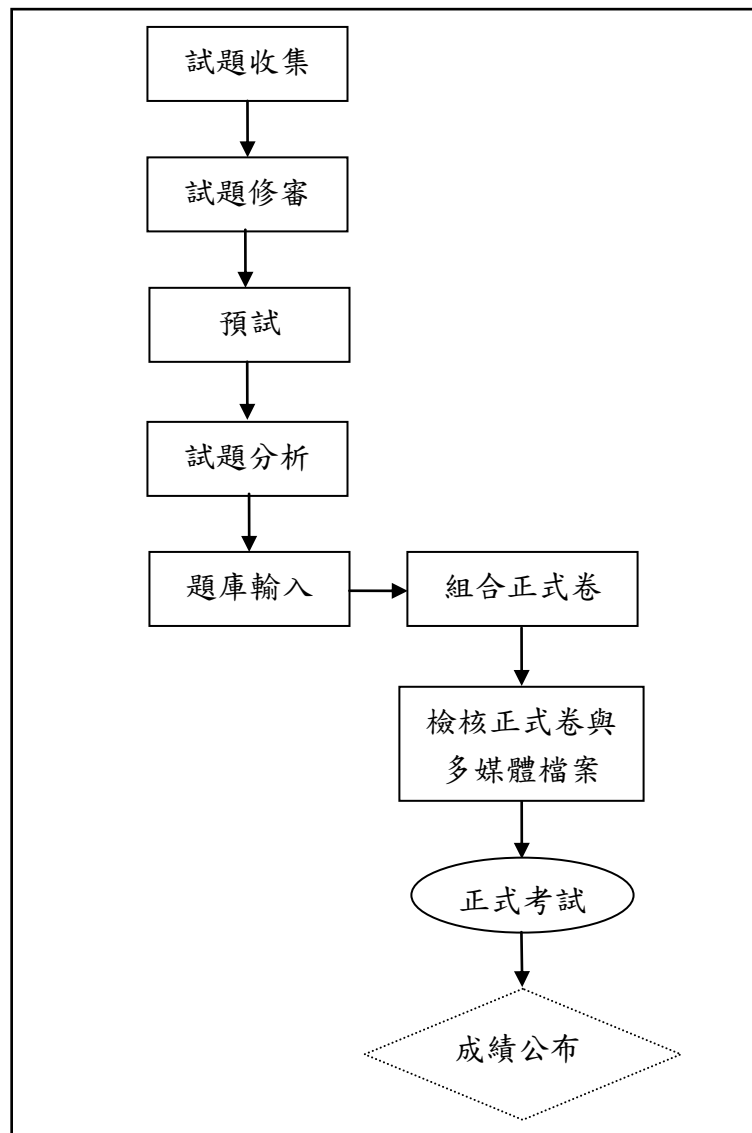


圖 1 正式考試標準化流程

(一) 試題收集

2013 年聽力測驗於 3 月舉辦入門基礎級、進階高階級與流利精通級三等的命題研習，參與研習教師共計 38 位，實際參與命題工作教師共 31 位。命題教師

每四個月寄回試題，由聽力測驗研發人員(以下簡稱研發人員)進行初步審題工作並與命題教師進行試題討論與修改。本年度命題回收數量為入門基礎級 925 題、進階高階級 96 題、流利精通級 411 題，共計 1,432 題。

(二) 試題修審

1. 會內初審

由本會非聽力測驗之研發人員進行第一階段會內初審，並回覆審題意見，接著再由研發人員根據審題意見修改試題，完成第一階段的審題工作。流利精通級試題在此階段除了進行會內初審外，還需進行「試考」，試考成員由非聽力測驗之研發人員組成，在模擬受測者考試環境下進行測驗，測驗結束以後，研發人員根據試考結果，逐一審查試題的編寫是否得當並進行修改，完成第一階段的審題工作。

2. 專家學者外審

由語言教學專家及資深華語教師，個別針對第一階段修改後的試題，進行第二階段之審題工作。在此階段中，專家學者主要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等提出審查意見。研發人員再依據審題意見修改試題，完成第二階段的審題工作。

3. 會內複審

針對第二階段專家學者外審修改後試題，由本會研究發展組組長進行複審，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、答案分布的比例是否平均等。最後，再由研發人員根據審查意見修改試題，完成第三階段的審題工作。

4. 製作試題相關多媒體檔案

完成上述試題修審三步驟後，接著製作試題相關多媒體檔案，包含委託專業錄音員至錄音室錄音、製作試題音檔及入門基礎級繪圖定稿。

(三) 預試

所有試題皆須經由預試提供評估試題品質之量化數據。本年度預試樣本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

由統計分析人員就預試之受測者反應進行試題分析，並以 IRT 做為分析取向。由於聽力測驗中各試題作答反應為非對即錯，屬於二元計分的試題 (dichotomous items)，故採取 IRT 之 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。評估標準之設定如下：凡試題之 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。除了提供試題的 Infit MNSQ、Infit ZSTD 指標外，尚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，以便研發人員能更全面地評估試題品質。

(五) 題庫輸入

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，本年度新增聽力測驗題庫數量為，入門基礎級 119 題、進階高階級 226 題、及流利精通級 155 題，共計 500 題。

(六) 組成正式卷

各等級測驗依照固定比例分配各難度區間之題數，再自本測驗題庫中挑選難度參數符合試卷難度區間要求的試題進行正式卷組卷。2012 年本會將基礎級試

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。

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，2014)⁶，再以此難度量尺為基準向下研發入門級試題，華語文聽力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、考點是否重複、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%	44.0%	-	-	-	-
-2.99~-2.50	14.0%	18.0%	-	-	-	-
-2.49~-2.00	8.0%	2.0%	-	-	-	-
-1.99~-1.50	4.0%	16.0%	-	-	-	-
-1.49~-1.00	18.0%	14.0%	40.0%	36.0%	-	-
-0.99~-0.50	2.0%	2.0%	2.0%	14.0%	-	2.0%
-0.49~-0.00	8.0%	2.0%	22.0%	14.0%	-	2.0%
0.01~0.50	2.0%	-	4.0%	4.0%	4.0%	4.0%
0.51~1.00	-	-	8.0%	16.0%	10.0%	18.0%
1.01~1.50	2.0%	2.0%	12.0%	4.0%	36.0%	22.0%
1.51~2.00	-	-	12.0%	8.0%	36.0%	32.0%
2.01~2.50	-	-	-	4.0%	10.0%	16.0%
2.51~3.00	-	-	-	-	4.0%	4.0%
>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

將步驟 6 正式卷中所有試題上傳至聽力測驗考試系統，由研發人員檢查試題圖片是否清楚、音檔音質是否清晰、試題文字是否無誤以及答案是否正確。確認無誤後，再由本會非聽力測驗研發人員進行與上述相同步驟之交叉檢核，經確認無誤後，即完成考試系統測試。

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。

⁶ 基礎級測驗試題連結至進階、高階與流利級測驗難度量尺之垂直等化研究為聽力、閱讀測驗同時進行。惟因篇幅有限，主要提及閱讀測驗，聽力測驗操作方式與閱讀測驗相同。

四、 測驗評估

一份測驗是否能夠發揮效用，確切地測量到欲測量的潛在能力(或潛在特質)，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 年華語文聽力測驗正式考試之信效度，來評估本測驗於此年度之測驗效能。本年度正式考試分別於 5 月 4 日、11 月 2 日舉行。五月份之到考人數分別為：入門基礎級 401 名、進階高階級 903 名及流利級 349 名。十一月份到考人數則分別為：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之測量精準度高。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 2013 年「華語文聽力測驗」之信度。

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。在 CTT 中，用來評估測驗信度的指標主要可分為四類：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，是指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。

以下在 CTT 取向採內部一致性信度，而 IRT 取向則採用測驗訊息量來評估聽力測驗信度。

1. 內部一致性信度

由於聽力測驗為受測者於同一時間接受同一份測驗，屬於僅施測一次之測驗形式，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指標的有：折半信度、Cronbach's α 係數及庫李 (Kuder-Richardson) 20 號或 21 號公式。以下將分析 2013 年華語文聽力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。

表 6 為各等級測驗於兩次正式考試中之內部一致性信度。由表 6 可知，在第一次正式考試中，各等級測驗信度係數均介於 .84 至 .90 之間；在第二次正式考試中，各等級測驗信度係數也介於 .88 至 .91 之間。由此顯示 2013 年華語文聽力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	.84	.88	.90
II	.90	.88	.91

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 40 至 42 分最大，數值為 9.16，測量誤差為 0.33，而在入門級與基礎門檻的測驗訊息量分別為 9.16 與 6.80，測量誤差分別為 0.33 與 0.38，信度係數分別為.89 與.85；第二次正式考試之測驗訊息量同樣以量尺分數 41 分為最大，數值為 9.87，測量誤差為 0.32，在入門級與基礎級門檻的測驗訊息量分別為 9.87 與 6.61，測量誤差分別為 0.32 與 0.39，信度係數分別為.94 與.90。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 9.89 與 7.58，以及 9.87 及 7.80，測量誤差均為 0.32 與 0.36，信度係數皆為.91 與.88。而流利級門檻在兩次正式考試的測驗訊息量則分別為 10.72 及 11.15，測量誤差為 0.31 及 0.30，信度係數為.94 及.93。

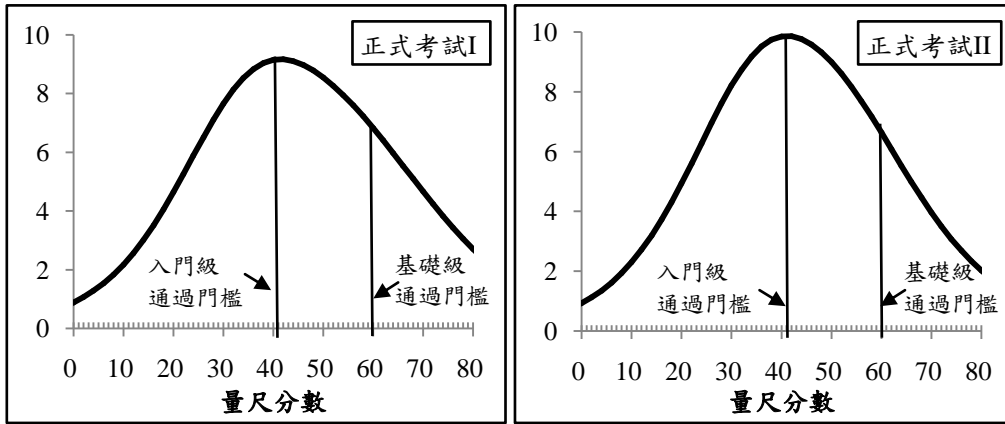


圖 2 入門基礎級測驗訊息量

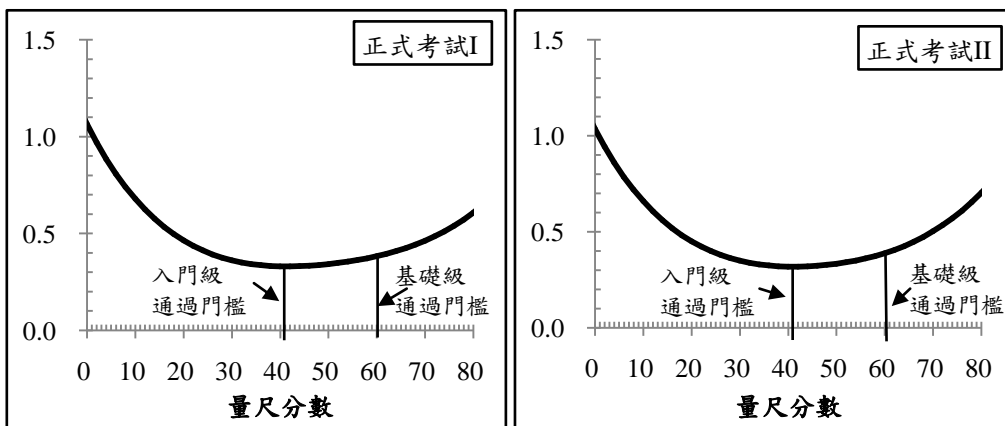


圖 3 入門基礎級測驗測量誤差

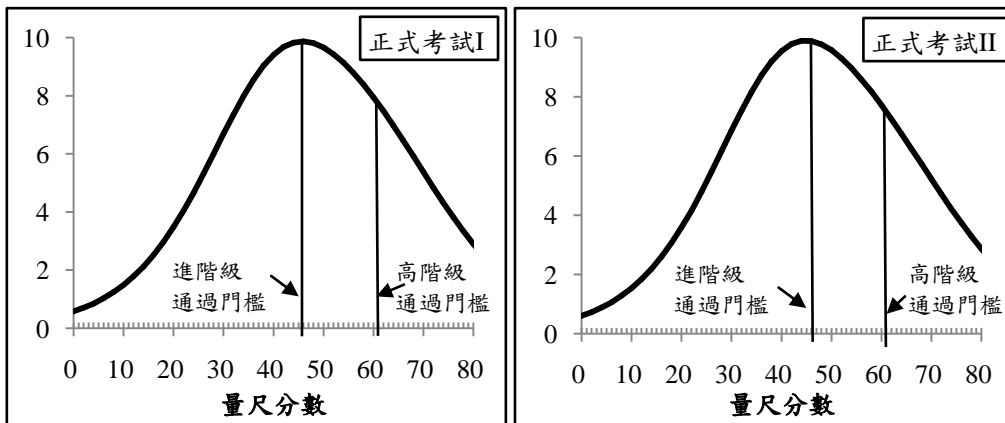


圖 4 進階高階級測驗訊息量

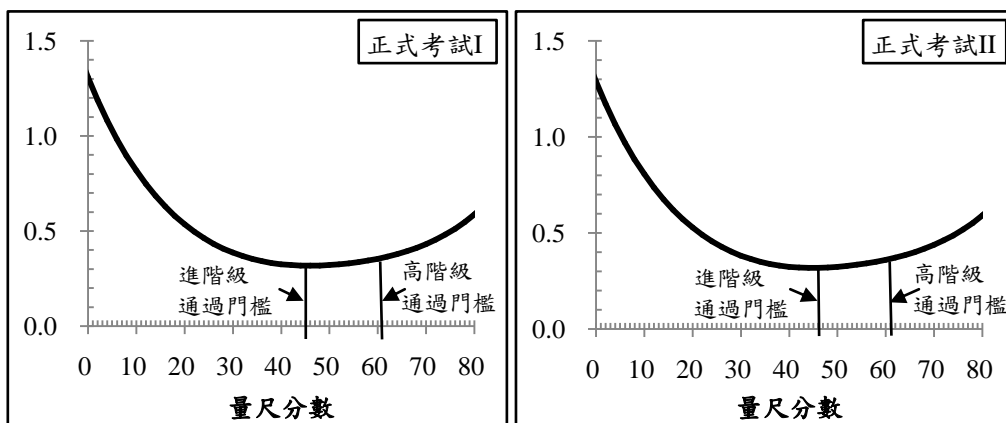


圖 5 進階高階級測驗測量誤差

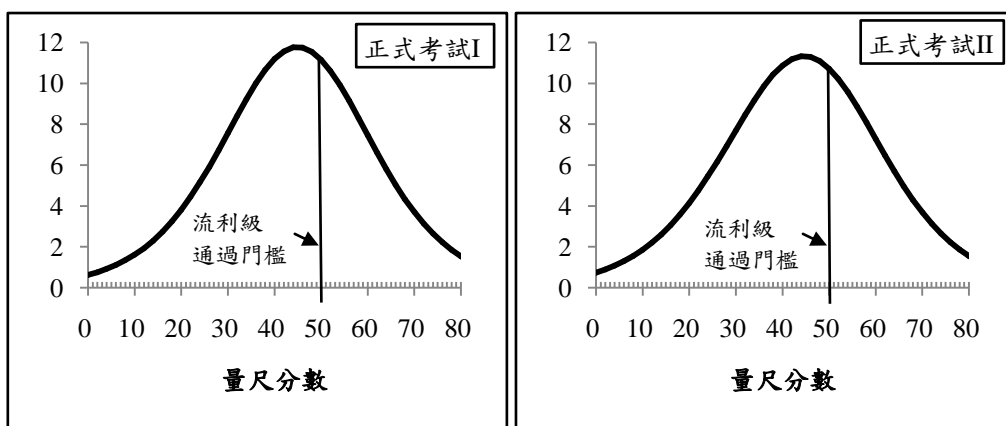


圖 6 流利級測驗訊息量

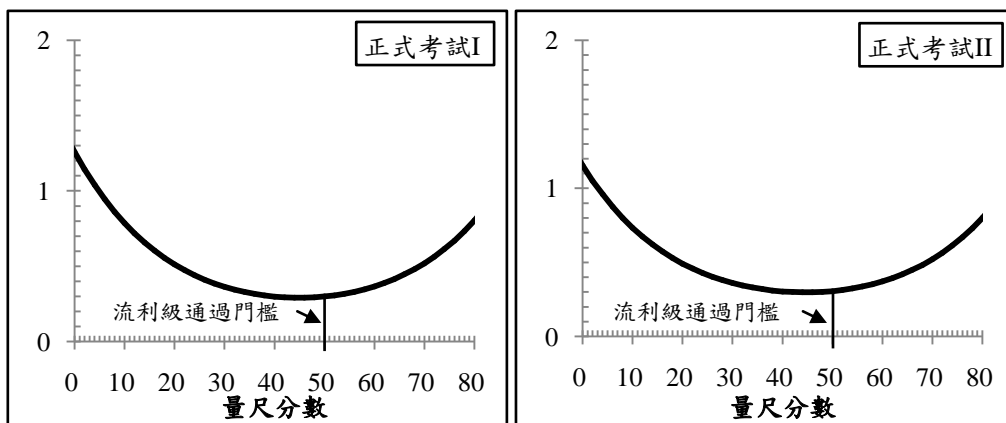


圖 7 流利級測驗測量誤差

(二) 效度

測驗效度指的是測驗是否能測量到欲測量的能力。由於欲測量的能力無法直接觀察，因此，須藉由受測者在試題上的作答反應或行為來間接推估。常用來驗證測驗效度的證據主要可分為三大類：1.內容效度(content validity)，即測驗內容的相關證據；2.建構效度(construct validity)，即關於測驗架構的證據；3.效標效

度(criterion validity)，即測驗結果預測力的相關證據。

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測驗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。此評估方式一般由該能力領域之學科專家進行，經由學科專家評估試題內容的方式，稱為專家效度。

建構效度相關證據，主要在評估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。古典測驗理論常採取因素分析，而 IRT 取向則會採用試題分析提供此部分證據。

效標效度的相關證據則可透過收集受測者之相關資料與測驗結果進行分析獲得。以下將從專家效度、試題分析及因素分析之驗證性因素分析(confirmatory factor analysis)來說明 2013 年「華語文聽力測驗」之內容效度及建構效度。

1. 內容效度

本會於發展聽力測驗之初，集結華語相關領域專家共同擬訂聽力測驗各等級測驗目標，並依據各等級目標訂立雙向細目表(如表 7)，命題教師經由培訓後，亦根據此雙向細目表及相關文件進行命題工作。接著由測驗研發人員回收各等級測驗試題，經正式測驗標準化流程(如圖 1)中試題修審步驟之第二階段：專家學者外審，審查各等級測驗試題內容是否符合該等級測驗通過者之能力描述(如表 1)、詞彙使用之適切性、以及題目(包含題幹、選項以及提問描述)表達之適當性，若不符合，則進行標註並寫明理由。測驗研發人員回收並整理專家學者外審意見後，逐一檢視並修改不符合之試題，若未標註則表示專家們對該等級試題內容具有一定共識。

表 7 雙向細目表

分項能力	題型 測驗等級	看圖	問答	對話	對話		段落	
		回答	理解	理解	進階	流利	進階	流利
		入門 基礎	入門 基礎	入門 基礎	進階 高階	流利	進階 高階	流利
整體性的聽力理解		V	V					
了解母語者間的交談			V	V	V	V		
聽宣布及說明指導			V	V			V	V
聆聽語音媒體與錄音					V	V	V	
作為現場觀眾一員的聽力表現							V	V

2. 建構效度

(1) 試題分析

由於聽力測驗是依據 IRT 試題難度參數進行組卷，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到相同潛在特質，也就是受測者回答試題皆仰賴單一特質。在此一前提下，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 模式進行資料分析，結果詳述如下。

整體而言，在試題難度的分布上，由於聽力測驗在組合正式卷時採用了固定比例來分配各難度區間之題數(詳見表 5)，因此，可以觀察到各等級兩次正式考試的試題難度估計平均值之絕對值差異約在 0.2 以內，而標準差的絕對值差異小於 0.2(如表 8)，顯示各等級之兩次正式考試試題難度分布相當相近。

在受測者能力分布方面(詳見表 9)，各等級測驗在兩次正式考試中，其受測者平均能力的絕對值差異約在 0.2 以內，而各等級受測者平均能力估計標準差的絕對值差異則小於 0.3。此亦表示各等級測驗在兩次正式考試中，受測者能力分布大致相近。

在入門基礎級兩次正式考試中，受測者平均能力估計值分別為-0.79與-0.59，50題試題的平均難度估計值分別為-2.41以及-2.62。以上結果顯示，兩次正式考試受測者平均能力與試題平均難度絕對值差異分別為1.6及2.0 logits，試題相對於受測者能力而言偏易；同樣的現象可由圖8受測者與試題分布圖(person-item map)觀察到，將受測者能力估計值與試題難度估計值放在一起(受測者能力置於圖的左方、試題難度置於圖的右方)，自上而下，將受測者能力由高到低、試題難度由難到易排列，發現試題難度涵蓋範圍與受測者能力分布範圍差距較大，受測者能力分布範圍略高於試題難度範圍，即測驗的試題相對於受測者來說較為簡單。雖然兩次正式考試試題平均難度與受測者平均能力差異較大，然而，所有試題難度範圍仍能涵蓋超過一半受測者能力範圍，例如，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83%與54%。而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，長時間沉浸於目標語環境中生活與學習，其學習成效較高，因此可以在較短時間內具備入門基礎級的聽力能力表現。

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	-2.41	-2.62	-0.14	-0.18	1.48	1.42
標準差	1.34	1.14	1.09	1.11	0.52	0.69
最大值	1.04	1.01	1.90	2.09	2.70	2.80
最小值	-3.88	-3.88	-1.38	-1.43	0.04	-0.69

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	-0.79	-0.59	0.47	0.68	2.22	2.13
標準差	1.04	1.32	1.05	1.09	1.24	1.25
最大值	3.71	3.32	3.60	5.60	6.72	6.73
最小值	-3.16	-3.72	-2.99	-2.19	-0.43	-0.56

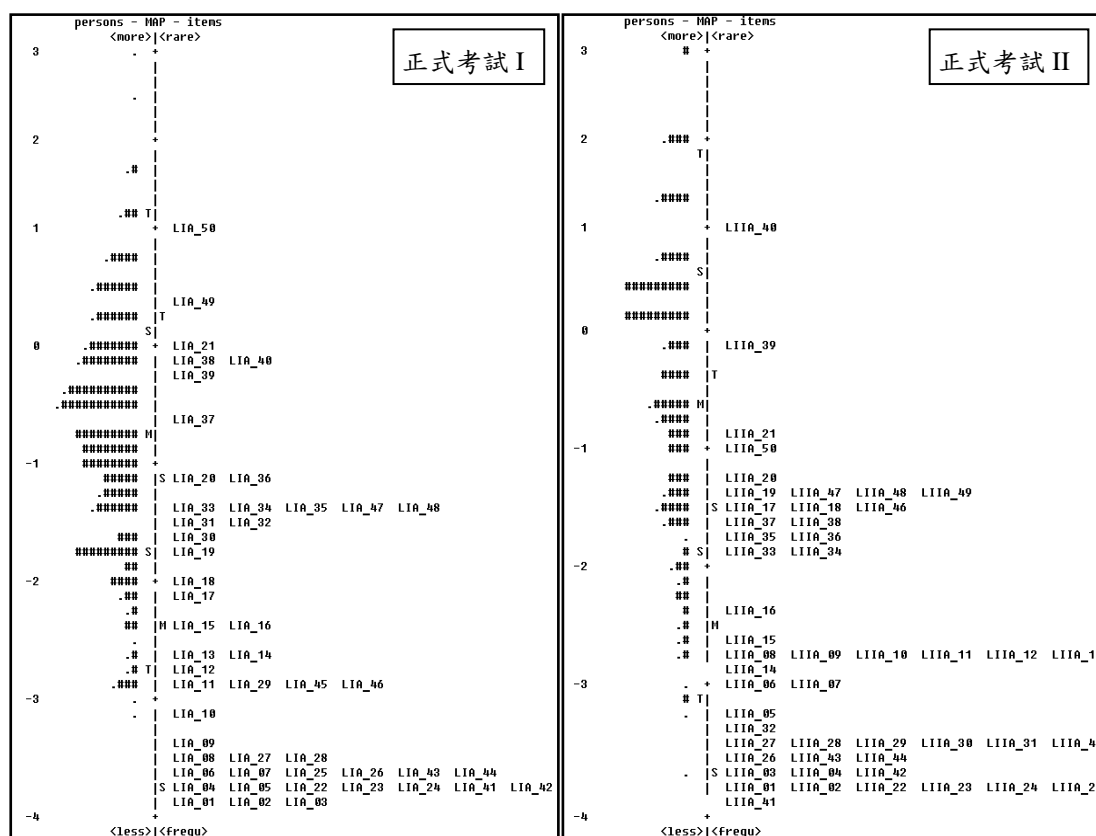


圖 8 入門基礎級測驗受測者與試題分布圖

註：“#”表示三位受測者；“.”表示一位受測者；LIA_01~LIA_50 和 LIIA_01~LIIA_50 為試題編號。

在進階高階級與流利級的兩次正式考試中，其結果與入門基礎級測驗相似，即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小，乃介於 0.6 至 0.9 logit 之間。由圖 9 和圖 10 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多數受測者能力分布範圍，進階高階級兩次正式考試試題平均難度 95% 信賴區間涵蓋的受測者能力分布分別為 93% 與 90%；流利級則為 65% 與 70%。

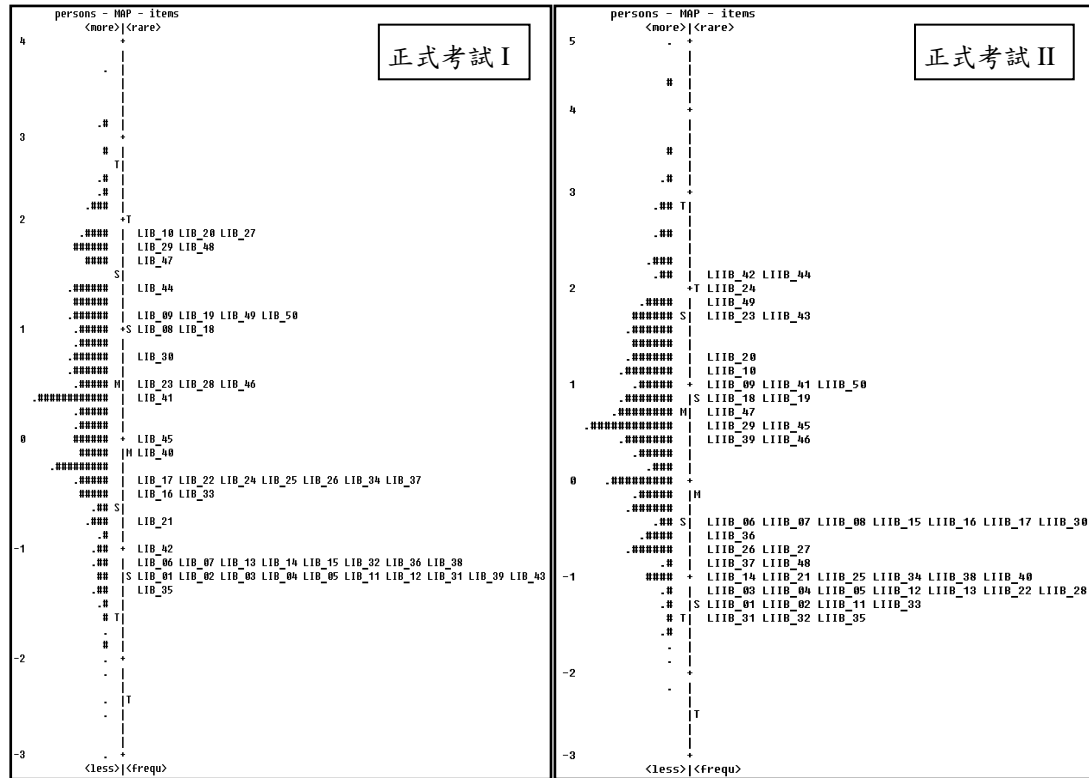


圖 9 進階高階級測驗受測者與試題分布圖

註：“#”分別表示兩次考試六位和五位受測者；“.”表示一位受測者；LIB_01~LIB_50 和 LIIB_01~LIIB_50 為試題編號。

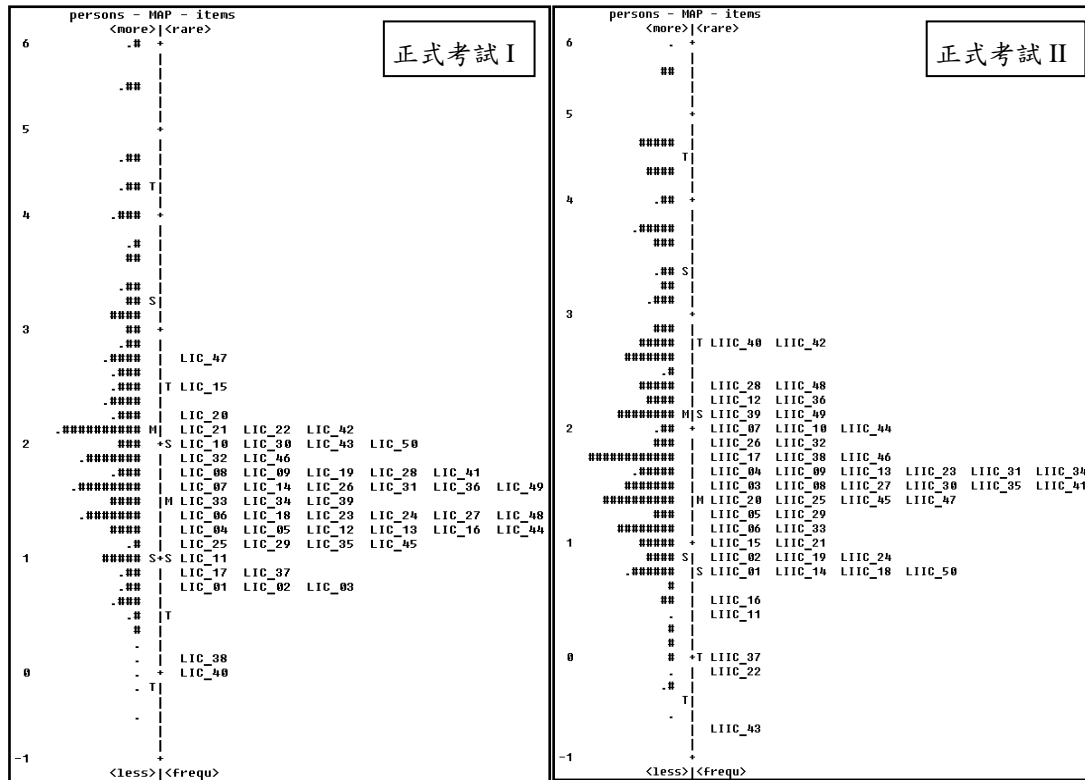


圖 10 流利級測驗受測者與試題分布圖

註：“#”分別表示兩次考試三位和二位受測者；“.”表示一位受測者；LIC_01~LIC_50 和 LIIC_01~LIIC_50 為試題編號。

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，分別有六道及 14 道試題不適配；進階高階級兩次測驗，50 道試題分別有三道、兩道試題不適配；而流利級兩次測驗各 50 道試題中，沒有試題不適配。除了入門基礎級第二次正式考試適配率稍低(72%)外，其餘各等級測驗正式考試的試題適配率介於 88% 至 100% 之間，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(此為聽力能力)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，試題的猜測力或鑑別度，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使得作答未符合其能力表現時，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條件下，高能力之受測者卻答錯較簡單題目。

由於本測驗正式卷的試題皆經過預試，且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	44	36	47	48	50	50
不適配題數	6	14	3	2	0	0
試題適配率 (%)	88	72	94	96	100	100

針對上述結果，進一步進行入門基礎級與進階高階級測驗不適配試題之選項分析，結果如下所述。表 11 和表 12 列出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、以及點二系列相關係數。

由表 11 可發現，入門基礎級兩次正式測驗不適配試題中，除了 LIA_50 外，各題選答正確選項人數都是最多的；而除了 LIA_49 和 LIIA_49 外，各題正確選項平均能力估計值也都是最高的，表示選答該選項的受測者平均能力高於選答錯誤選項者，如模式所預期，能力越高的受測者答對試題的機率越高。而雖然 LIA_50 選答正確選項考數低於誘答，但正確選項考生平均能力值仍為最高；LIA_49 和 LIIA_49 正確選項考生平均能力值雖然低於漏答考生，但兩道試題各僅有一名和兩名考生漏答。此外，所有不適配試題中正確選項的點二系列值都較高且皆為正值，相關係數介於.11 至.55，表示具有幅合效度(convergent validity)，即能力越高者答對試題的機率越高；錯誤選項值較低且幾乎為負值，表示具有區辨效度(discriminate validity)，即可有效區分能力高與低的受測者(Linacre, 2010)。其中，在二次正式測驗中，僅有一題(LIIA_35)的正確選項點二系列相關較低，未達.20，從圖 11 試題特徵曲線(item characteristic curve，簡稱 ICC)來看，能力值與試題難度差異介於-2.0~-1.0 的受測者答對此題機率高於模式預期；而能力值與試題難度差異介於 1.0~2.5 與 4.0 左右的受測者答對此題機率則低於模式預期，使得相關係數較低，也因此造成觀察資料與模式不適配。

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	LIA_05	-3.751	0.200	B	0	12	-1.61	0.25	-.13
				A	0	47	-1.42	0.13	-.18
				C	1	340	-0.67	0.06	.23
				M	0	2	-0.71	0.16	.03
	LIA_10	-3.173	0.164	C	0	10	-2.29	0.23	-.25
				A	0	2	-2.12	0.82	-.09
				B	1	388	-0.74	0.05	.27
				M	0	1	-1.93	--	-.05
	LIA_22	-3.792	0.203	B	0	11	-2.26	0.19	-.26
				C	0	48	-1.34	0.14	-.17
				A	1	336	-0.65	0.05	.27
				M	0	6	-1.40	0.37	-.06
	LIA_45	-2.910	0.151	A	0	20	-1.65	0.16	-.17
				C	0	73	-1.19	0.11	-.13
				B	1	307	-0.64	0.06	.21
				M	0	1	-1.17	--	-.01
	LIA_49	0.431	0.123	C	0	96	-1.42	0.08	-.30
				B	0	82	-0.94	0.10	-.01
				A	1	222	-0.46*	0.07	.26
				M	0	1	-0.38	--	.03
LIA_50	1.038	0.141	B	0	205	-1.07	0.06	-.20	
			C	0	66	-1.03	0.09	-.05	
			A	1	128	-0.19	0.11	.25	
			M	0	2	-2.21	0.40	-.11	
LIIA_03	-3.731	0.251	C	0	1	-3.09	--	-.15	
			B	0	5	-2.12	0.22	-.16	
			A	1	249	-0.55	0.08	.21	
			B	0	9	-2.06	0.45	-.25	
LIIA_04	-3.731	0.249	C	0	23	-1.13	0.25	-.09	
			A	1	222	-0.46	0.09	.22	
			M	0	1	-3.72	--	-.20	
			A	0	3	-2.49	0.46	-.19	
LIIA_07	-2.985	0.200	C	0	47	-1.30	0.13	-.19	
			B	1	202	-0.37	0.09	.23	
			M	0	3	-2.21	0.76	-.15	
			C	0	2	-2.88	0.00	-.19	
LIIA_09	-2.903	0.196	B	0	15	-2.02	0.27	-.30	
			A	1	237	-0.47	0.08	.35	
			M	0	1	-3.72	--	-.20	

表 11 入門基礎級測驗不適配試題選項描述性摘要表(續)

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	LIIA_10	-2.903	0.196	A	0	10	-2.39	0.31	-.34
				C	0	41	-1.19	0.17	-.16
				B	1	201	-0.34	0.09	.31
				M	0	3	-2.86	0.44	-.23
	LIIA_23	-3.846	0.261	C	0	9	-2.08	0.29	-.24
				B	0	21	-1.88	0.19	-.30
				A	1	216	-0.36	0.09	.40
				M	0	9	-1.54	0.37	-.13
	LIIA_25	-3.830	0.260	C	0	6	-2.83	0.13	-.34
				B	0	22	-1.67	0.22	-.25
				A	1	218	-0.39	0.08	.39
				M	0	9	-1.27	0.42	-.08
	LIIA_28	-3.470	0.230	A	0	4	-2.88	0.11	-.27
				C	0	1	-2.88	--	-.13
				B	1	250	-0.54	0.08	.30
	LIIA_30	-3.442	0.228	A	0	25	-1.97	0.19	-.36
				C	0	22	-1.59	0.22	-.21
				B	1	208	-0.32	0.09	.43
	LIIA_31	-3.438	0.228	B	0	39	-2.00	0.13	-.49
				A	0	21	-1.53	0.24	-.20
C				1	192	-0.19	0.08	.55	
M				0	3	-1.50	0.93	-.08	
LIIA_33	-1.912	0.160	A	0	42	-1.41	0.16	-.24	
			C	0	45	-1.02	0.17	-.10	
			B	1	162	-0.25	0.10	.27	
			M	0	6	-0.86	0.34	.01	
LIIA_35	-1.782	0.157	A	0	25	-1.06	0.23	-.09	
			C	0	83	-0.90	0.12	-.06	
			B	1	139	-0.33	0.12	.11	
			M	0	8	-0.45	0.43	.04	
LIIA_43	-3.683	0.247	C	0	1	-3.19	--	-.16	
			B	0	2	-2.52	0.36	-.15	
			A	1	252	-0.56	0.08	.21	
LIIA_49	-1.405	0.151	B	0	86	-1.19	0.12	-.27	
			C	0	61	-0.86	0.15	-.06	
			A	1	106	0.03*	0.12	.31	
			M	0	2	0.30	0.45	.07	

註：M 表示漏答；"--"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；"*"表示選答正確選項受測者平均能力估計值非最高。

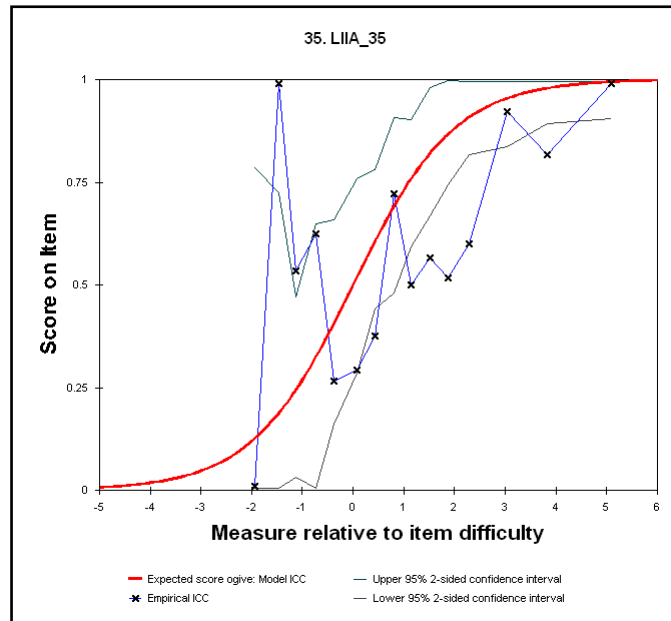


圖 11 LIIA_35 試題特徵曲線

進階高階級測驗選項分析結果如表 12 所示，兩次正式考試共有五道試題不適配，其中，LIB_43 和 LIB_02 選答正確選項人數較多，選答正確選項受測者平均能力值最高，且正確選項點二系列相關係數較高且為正值，表示具有幅合效度，即能力越高者答對試題的機率越高；錯誤選項值較低且為負值，表示具有區辨效度，即可有效區分能力高與低的受測者(Linacre, 2010)。其餘三題(LIB_27、LIB_48 和 LIIB_50)正答選項點二系列相關數值均低於.20，其可能原因為受測者作答情形與其能力不相符，以及正確選項與某一誘答選項受測者平均能力值較為接近所致。例如三道試題平均難度與受測者平均能力差異分別為 1.43、1.32 與 0.31 logits，依模式預期(如公式 1)，該三道試題對於受測者(如表 8)來說，平均答對機率分別為 19%、21%及 42%，然而，卻有 31%、32%及 70%受測者答對此三道試題，顯示答對此三道試題之受測者中有部分受測者其作答情形與能力不相符。同樣的，由三道試題之試題特徵曲線(圖 12 到圖 14)可觀察到，三道試題在部分能力較低的受測者，答對率高於模式預期；而部分能力較高的受測者，答對率低於預期。且三道試題中選答正確選項及某一誘答選項之受測者平均能力差異介於 0.10 至 0.28 之間，顯示試題無法有效區分出不同能力受測者。

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	LIB_27	1.901	0.086	C	0	124	0.05	0.09	-.15
				A	0	123	0.15	0.08	-.10
				B	0	362	0.49	0.05	.05
				D	1	279	0.77	0.07	.13
				M	0	15	0.22	0.20	-.02
	LIB_43	-1.225	0.090	B	0	25	-0.26	0.23	-.11
				A	0	32	-0.09	0.14	-.09
				C	0	213	0.02	0.07	-.20
				D	1	628	0.68	0.04	.26
				M	0	5	0.15	0.23	-.01
	LIB_48	1.794	0.084	B	0	255	0.08	0.06	-.21
				A	0	107	0.41	0.09	.00
				C	0	245	0.60	0.07	.10
				D	1	288	0.74	0.07	.11
				M	0	8	-0.03	0.20	-.04
II	LIIB_02	-1.245	0.105	A	0	21	-0.34	0.12	-.16
				C	0	61	-0.34	0.09	-.29
				D	0	135	-0.03	0.07	-.29
				B	1	517	1.04	0.04	.48
				M	0	8	0.05	0.15	-.05
	LIIB_50	0.990	0.082	C	0	23	0.08	0.21	-.09
				A	0	37	0.11	0.14	-.11
				B	0	159	0.66	0.09	.03
				D	1	516	0.76	0.05	.06
				M	0	7	0.04	0.44	-.05

註：M 表示漏答。

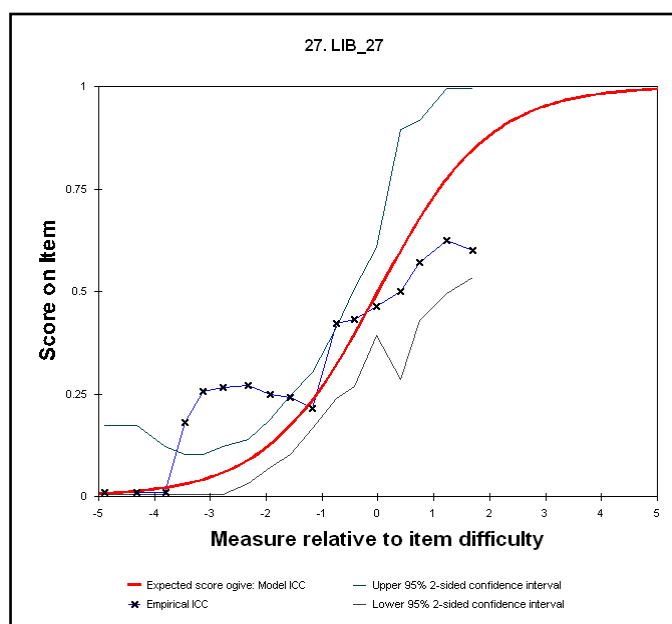


圖 12 LIB_27 試題特徵曲線

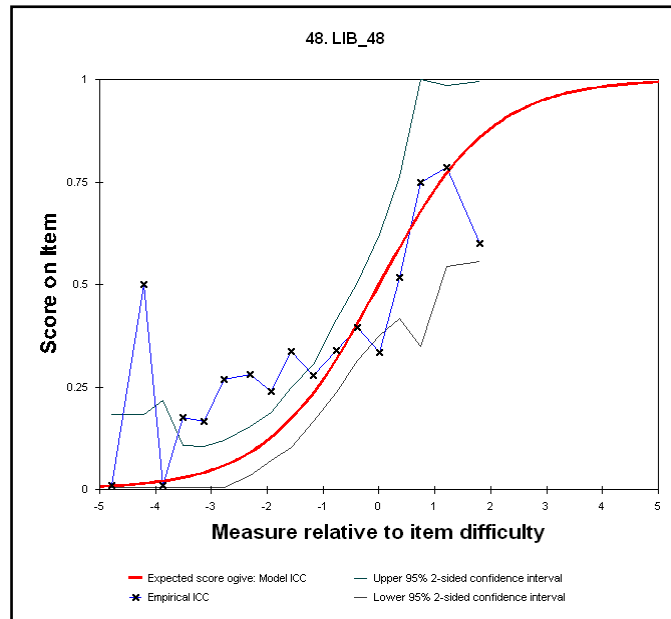


圖 13 LIB_48 試題特徵曲線

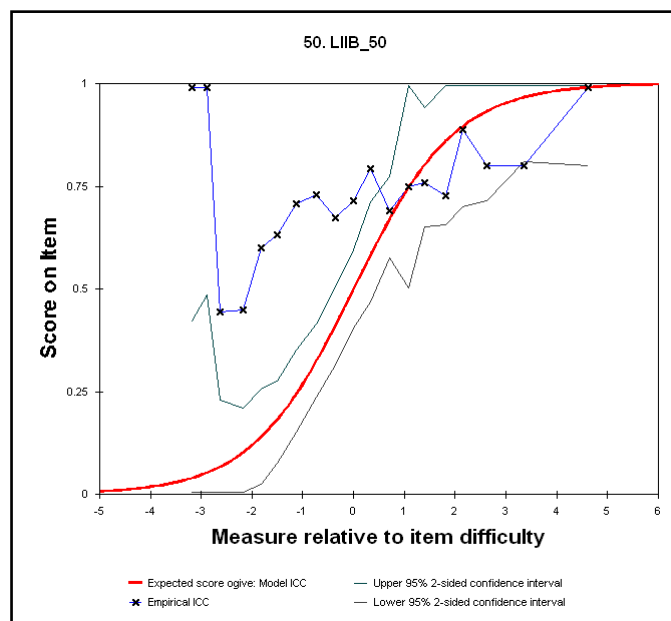


圖 14 LIIB_50 試題特徵曲線

(2) 驗證性因素分析

除了透過試題分析來評估聽力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。由於聽力測驗題型為單選題，答題反應皆為非對即錯的二分名義變項，故使用可分析二元變項和類別變項的 Mplus 分析軟體，來檢視聽力測驗入門基礎級、進階高階級，與流利級雙向細目中欲測得之聽力分項能力，是否可經由受測者

於試題的答題反應測知。每一等級之因素分析模型中，測量變數為每道試題，而潛在變數則為聽力分項能力，例如，本年度5月及11月入門基礎級測驗中，測量變數為50道試題，潛在變數為「整體性的聽力理解」與「了解母語使用者間的交談」。

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)，訂定出基本適配度的評估標準如下：(1)因素負荷量介於.50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，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(root mean square error of approximation；簡稱RMSEA)指標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，聽力測驗入門基礎級、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15至圖17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.09至.82之間，平均值為.49；因素負荷量標準誤為0.05至0.10；除LIA_39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.92。11月正式考試因素負荷量介於.12至.88之間，平均值為.59；因素負荷量標準誤為0.04至0.11；除LIIA_21、LIIA_35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.95。

進階高階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.77之間，平均值為.50；因素負荷量標準誤為0.03至0.05；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.95、.97及.91。11月正式考試因素負荷量介於.09至.71之間，平均值為.48；因素負荷量標準誤為0.03至0.07；誤差變異量皆為正值，除LIIB_50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.98、.97及.95。

流利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.25至.70之間，平均值為.53；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.88、.95及.93。11月正式考試因素負荷量介於.31至.84之間，平均值為.56；因素負荷量標準誤為0.05至0.08；因素負荷量統計考驗達顯著水準($p < .05$)；潛在變項間的相關係數為.52、.93及.63。

根據以上指標分析，可知聽力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48%至74%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的試題比例為10%與6%，進階高階級和流利級依序為14%、10%、4%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。

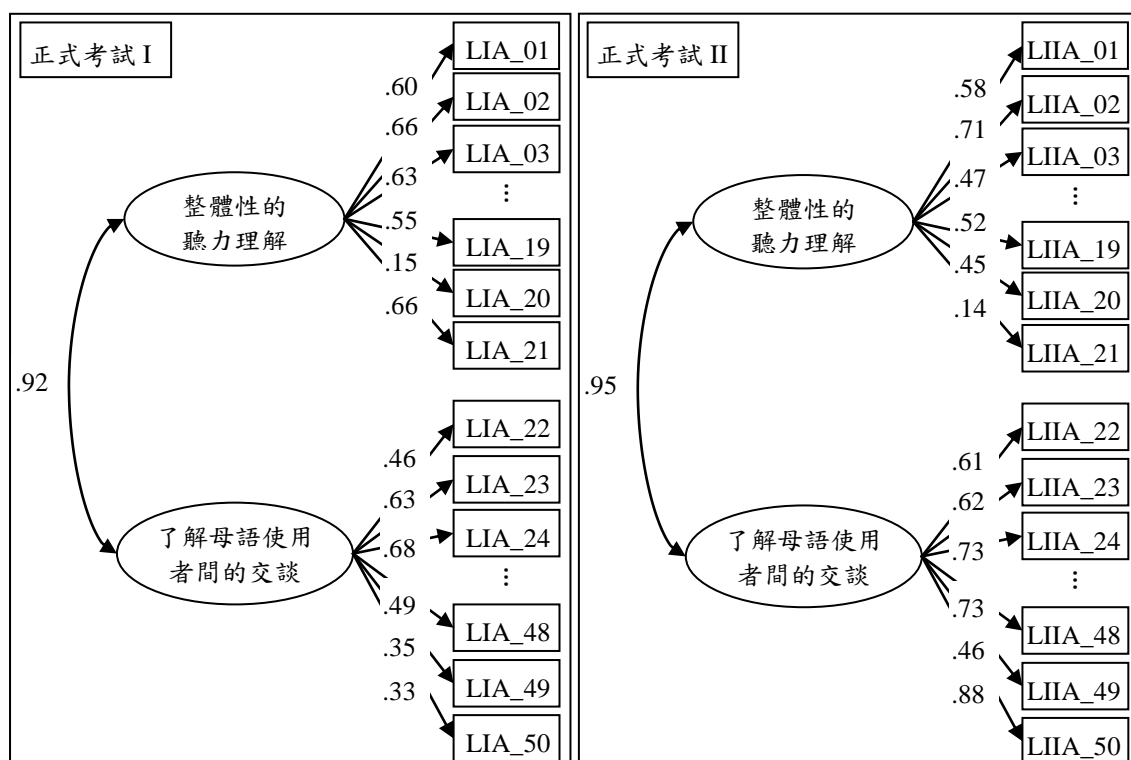


圖 15 入門基礎級測驗二因素驗證性因素分析

註：LIA_01-LIA_21、LIIA_01-LIIA_21所屬分項能力為「整體性的聽力理解」；LIA_22-LIA50、LIIA_22-LIIA50所屬分項能力為「了解母語使用者間的交談」。

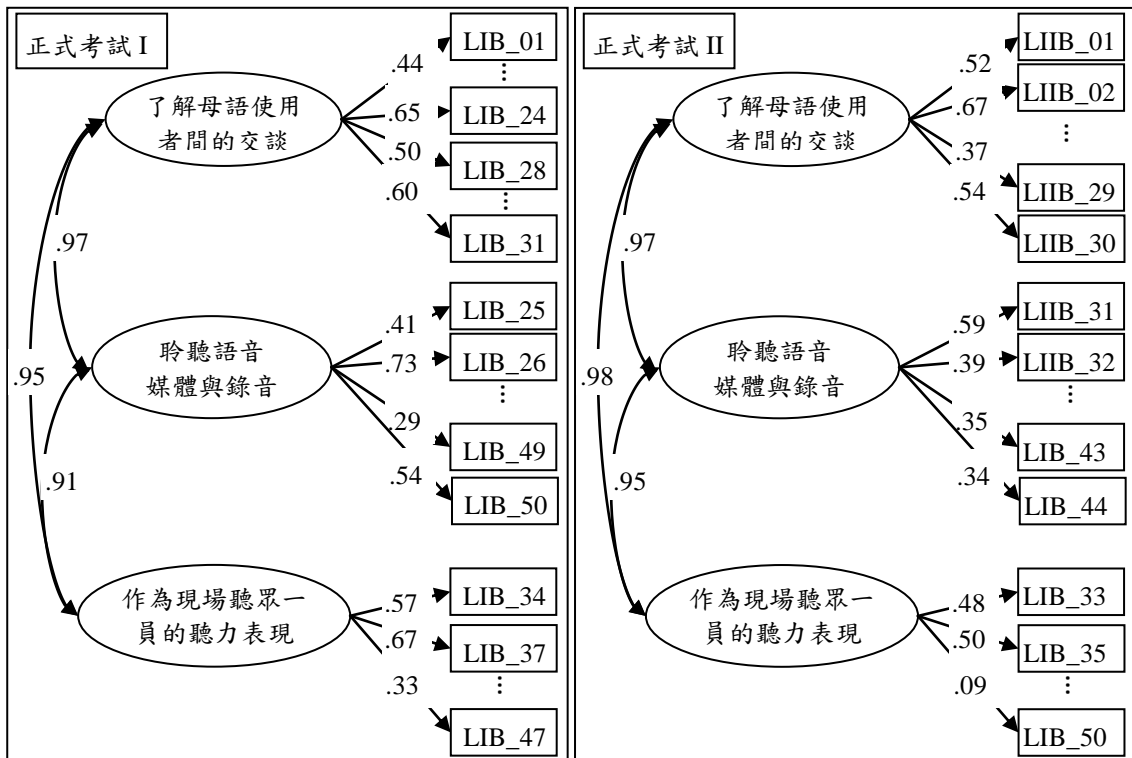


圖 16 進階高階級測驗三因素驗證性因素分析

註：LIB_01-LIB_24、LIB_28-LIB_31、LIIB_01-LIIB_30所屬分項能力為「了解母語使用者間的交談」；LIB_25-LIB_27、LIB_32、LIB_33、LIB_35、LIB_36、LIB_39-LIB_44、LIB_48-LIB_50、LIIB_31、LIIB_32、LIIB_34、LIIB_39-LIIB_44所屬分項能力為「聆聽語音媒體與錄音」；LIB_34、LIB_37、LIB_38、LIB_45-LIB_47、LIIB_33、LIIB_35-LIIB_38、LIIB_45-LIIB_50所屬分項能力為「作為現場聽眾一員的聽力表現」。

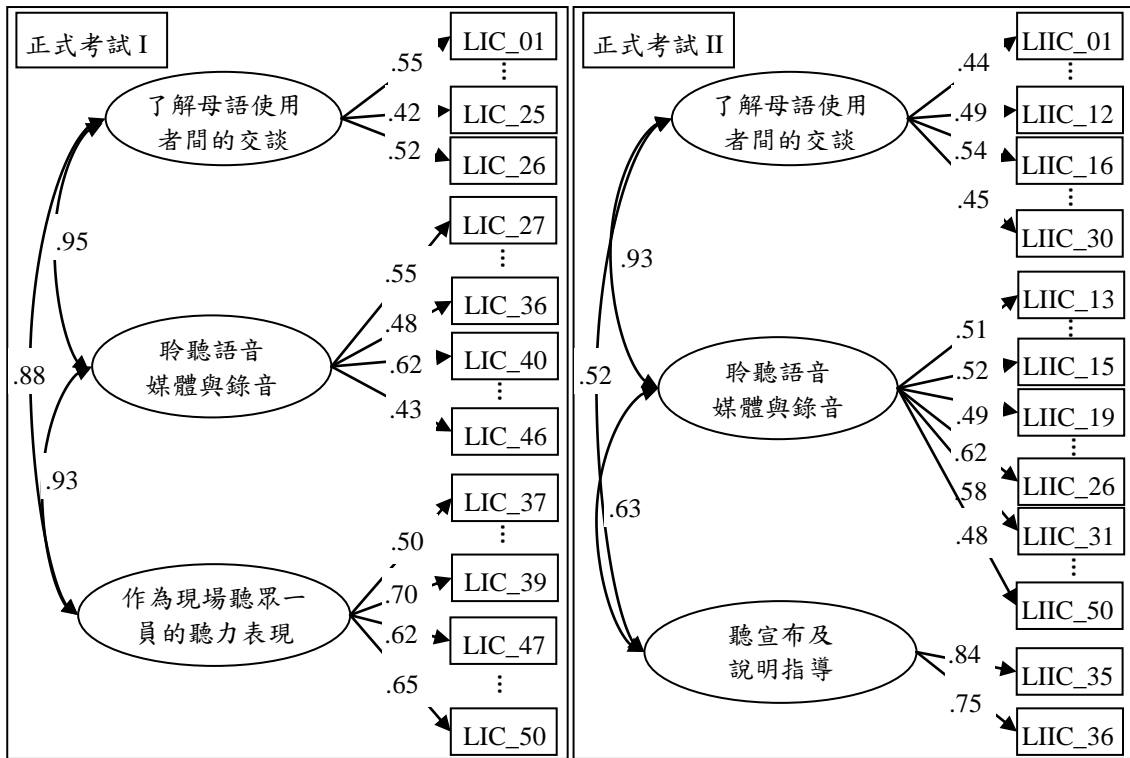


圖 17 流利級測驗三因素驗證性因素分析

註：LIC_01-LIC_26、LIIC_01-LIIC_12、LIIC_16-LIIC_18、LIIC_27-LIIC_30 所屬分項能力為「了解母語使用者間的交談」；LIC_27-LIC_36、LIC_40-LIC_46、LIIC_13-LIIC15、LIIC_19-LIIC26、LIIC_31-LIIC34、LIIC_37-LIIC_50 所屬分項能力為「聆聽語音媒體與錄音」；LIC_37-LIC_39、LIC_47-LIC_50 所屬分項能力為「作為現場聽眾一員的聽力表現」；LIIC_35、LIIC_36 所屬分項能力為「聽宣布及說明指導」。

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)來評估模式適配度，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 13 可發現，聽力測驗各等級 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，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，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，各等級聽力測驗具有建構效度，各測驗試題分別可測得不同的聽力分項能力，而各聽力分項能

力間的相關係數達到中度至高度相關，反映測得一致之聽力理解能力。

表 13 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	1.11	.02	.94	.94
	II	1.13	.03	.95	.95
進階高階級三因素模式	I	1.39	.02	.97	.96
	II	1.29	.02	.96	.96
流利級三因素模式	I	1.20	.02	.95	.95
	II	1.12	.02	.97	.97

五、 結論

2013 年華語文聽力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)；以及流利級(Level 5)。五級分別對應於 CEFR 之 A1 (Break-through)、A2 (Waystage)、B1 (Threshold)、B2 (Vantage) 與 C1 (Effective Operational Proficiency)。各等級通過門檻均經由標準設定(standard setting)程序而來。且自本年度起，以量尺分數(scale score)提供受測者正式考試測驗結果與等級通過與否之判斷依據，此方式可避免測驗通過標準受不同正式卷試題間具有不同難度而影響，使測驗更具客觀性與公平性。

本文除說明聽力測驗之能力描述、測驗題型題數及研發、施測與成績公布等標準化流程外，在通過門檻部分，亦針對量尺分數的由來進行說明，最後則透過分析 2013 年度全國性正式考試之信度與效度，來進行此年度聽力測驗整體性的測驗效能評估，以檢視其是否能夠發揮效用，確切地測量受測者的聽力能力。

在測驗信度分析方面，藉由內部一致性指標來表示試題間之相關性，以確認試題皆測量到相同潛在特質的程度為何，以及透過測驗訊息量來表徵測量精準度。而在測驗效度分析部分，首先，以專家審查及評估各等級試題適切性來確保試題品質、確認試題內容，因此，可作為測驗之內容效度的證據。接著，經由受測者作答反應分別進行試題分析及驗證性因素分析，以確認由受測者反應所建構出的測驗架構是否與測驗發展之初所制訂的相同，以提供測驗之建構效度證據。經由信度及效度分析結果，顯示各等級聽力測驗皆具有信度及效度，發揮了測驗效能。

由上述可知，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使用於測量受測者之聽力能力，故專家審查程序是相當重要的，為了提供更嚴謹、全面及量化的專家效度證據，本會未來將更具針對性地提供專家外審參考資料，以便專家更好地完成修審工作。比如設計審題問卷，請專家針對每一道試題進行詳細評分，再取平均數或中位數作為專家試題評估的量化結果。

六、文獻

- 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。
- 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.
- 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.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*
Copenhagen: Institute of Educational Research. (Expanded edition, 1980. Chi-
cago: The University of Chicago Press.)

附件 1 聽力測驗各等級標準設定會議問卷調查結果⁷

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目的說明得很清楚。	3.9
2. 會議帶領者對於 CEFR 的內容和架構說明得很清楚。	3.7
3. 我了解最低能力者在標準設定方法上的涵義。	3.8
4. 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。	3.6
5. 標準設定的練習活動有助於我了解如何進行任務。	3.8
6. 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。	3.8
7. 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
8. 試題的難度會影響我對試題的判斷。	3.1
9. 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。	3.2
10. 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. 會議帶領者對於試題難度參數的意涵解釋得很清楚。	3.9
2. 我了解最低能力者在標準設定方法上的涵義。	3.7
3. 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。	3.8
4. 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
5. 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。	3.8
6. 提供的試題難度資料會影響我對試題的判斷。	3.3
7. 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。	3.1
8. 我對於自己所設定的切截分數(cut score)有信心。	3.2

附件 1-3 流利精通級調查結果

問卷題目	平均數
1. 會議帶領者對於本次會議的目的說明得很清楚。	3.9
2. 會議帶領者對於 CEFR 的內容和架構說明得很清楚。	3.8
3. 我了解最低能力者在標準設定方法上的涵義。	3.8
4. 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。	3.7
5. 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
6. 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
7. 試題的難度會影響我對試題的判斷。	3.3
8. 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
9. 我對於自己所設定的切截分數(cut score)有信心。	3.5

⁷ 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 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	整體性的聽力理解	LIA_01	.60	0.08	.00
		LIA_02	.66	0.07	.00
		LIA_03	.63	0.08	.00
		LIA_04	.53	0.08	.00
		LIA_05	.38	0.08	.00
		LIA_06	.30	0.10	.00
		LIA_07	.74	0.06	.00
		LIA_08	.59	0.09	.00
		LIA_09	.50	0.08	.00
		LIA_10	.67	0.09	.00
		LIA_11	.45	0.09	.00
		LIA_12	.66	0.06	.00
		LIA_13	.50	0.07	.00
		LIA_14	.50	0.07	.00
		LIA_15	.48	0.07	.00
		LIA_16	.43	0.08	.00
		LIA_17	.66	0.06	.00
		LIA_18	.48	0.06	.00
		LIA_19	.55	0.06	.00
		LIA_20	.15	0.07	.04
		LIA_21	.66	0.06	.00
	LIA_22	.46	0.07	.00	
	LIA_23	.63	0.07	.00	
	LIA_24	.68	0.08	.00	
	LIA_25	.33	0.10	.00	
	LIA_26	.56	0.08	.00	
	LIA_27	.37	0.10	.00	
	LIA_28	.52	0.09	.00	
	LIA_29	.45	0.08	.00	
	LIA_30	.33	0.07	.00	
	LIA_31	.23	0.07	.00	
	LIA_32	.34	0.07	.00	
	LIA_33	.35	0.07	.00	
	LIA_34	.29	0.07	.00	
	LIA_35	.20	0.07	.00	
	LIA_36	.55	0.06	.00	
	LIA_37	.64	0.05	.00	
	LIA_38	.53	0.06	.00	
	LIA_39	.09	0.08	.28	
	LIA_40	.42	0.07	.00	
	LIA_41	.61	0.08	.00	
	LIA_42	.76	0.06	.00	
	LIA_43	.82	0.05	.00	
	LIA_44	.77	0.06	.00	
	LIA_45	.30	0.07	.00	
	LIA_46	.33	0.09	.00	
	LIA_47	.67	0.05	.00	
	LIA_48	.49	0.06	.00	
	LIA_49	.35	0.07	.00	
	LIA_50	.33	0.08	.00	
	了解母語使用者間的交談				

正式考試	分項能力	試題編號	因素負荷量	標準誤	P 值
II	整體性的聽力理解	LIIA_01	.58	0.07	.00
		LIIA_02	.71	0.08	.00
		LIIA_03	.47	0.06	.00
		LIIA_04	.42	0.10	.00
		LIIA_05	.59	0.08	.00
		LIIA_06	.58	0.09	.00
		LIIA_07	.38	0.08	.00
		LIIA_08	.41	0.11	.00
		LIIA_09	.69	0.09	.00
		LIIA_10	.48	0.08	.00
		LIIA_11	.61	0.08	.00
		LIIA_12	.72	0.06	.00
		LIIA_13	.22	0.11	.04
		LIIA_14	.63	0.08	.00
		LIIA_15	.78	0.06	.00
		LIIA_16	.63	0.07	.00
		LIIA_17	.85	0.05	.00
		LIIA_18	.46	0.08	.00
		LIIA_19	.52	0.07	.00
		LIIA_20	.45	0.07	.00
		LIIA_21	.14	0.09	.10
	LIIA_22	.61	0.10	.00	
	LIIA_23	.62	0.07	.00	
	LIIA_24	.73	0.07	.00	
	LIIA_25	.56	0.08	.00	
	LIIA_26	.84	0.05	.00	
	LIIA_27	.38	0.10	.00	
	LIIA_28	.80	0.06	.00	
	LIIA_29	.65	0.09	.00	
	LIIA_30	.64	0.07	.00	
	LIIA_31	.75	0.05	.00	
	LIIA_32	.68	0.07	.00	
	LIIA_33	.36	0.08	.00	
	LIIA_34	.44	0.08	.00	
	LIIA_35	.12	0.09	.16	
	LIIA_36	.72	0.06	.00	
	LIIA_37	.83	0.05	.00	
	LIIA_38	.54	0.07	.00	
	LIIA_39	.55	0.07	.00	
	LIIA_40	.72	0.05	.00	
	LIIA_41	.66	0.07	.00	
	LIIA_42	.54	0.08	.00	
	LIIA_43	.63	0.08	.00	
	LIIA_44	.70	0.06	.00	
	LIIA_45	.83	0.05	.00	
	LIIA_46	.37	0.08	.00	
	LIIA_47	.74	0.06	.00	
	LIIA_48	.73	0.05	.00	
	LIIA_49	.46	0.07	.00	
	LIIA_50	.88	0.04	.00	
	了解母語使用者間的交談				

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	了解母語使用者間的交談	LIB_1	.44	0.05	.00
		LIB_2	.71	0.03	.00
		LIB_3	.59	0.04	.00
		LIB_4	.77	0.03	.00
		LIB_5	.72	0.03	.00
		LIB_6	.41	0.05	.00
		LIB_7	.71	0.03	.00
		LIB_8	.48	0.04	.00
		LIB_9	.29	0.04	.00
		LIB_10	.49	0.05	.00
		LIB_11	.52	0.04	.00
		LIB_12	.71	0.03	.00
		LIB_13	.54	0.04	.00
		LIB_14	.19	0.05	.00
		LIB_15	.55	0.04	.00
		LIB_16	.49	0.04	.00
		LIB_17	.61	0.03	.00
		LIB_18	.19	0.04	.00
		LIB_19	.34	0.04	.00
		LIB_20	.30	0.04	.00
		LIB_21	.45	0.04	.00
		LIB_22	.49	0.04	.00
		LIB_23	.57	0.04	.00
		LIB_24	.65	0.03	.00
		LIB_28	.50	0.04	.00
		LIB_29	.41	0.04	.00
		LIB_30	.36	0.04	.00
		LIB_31	.60	0.04	.00
		LIB_25	.41	0.04	.00
		LIB_26	.73	0.03	.00
		LIB_27	.17	0.05	.00
	LIB_32	.60	0.04	.00	
	LIB_33	.48	0.04	.00	
	LIB_35	.55	0.04	.00	
	LIB_36	.74	0.03	.00	
	LIB_39	.71	0.04	.00	
	LIB_40	.35	0.04	.00	
	LIB_41	.52	0.04	.00	
	LIB_42	.57	0.04	.00	
	LIB_43	.37	0.04	.00	
	LIB_44	.21	0.05	.00	
	LIB_48	.15	0.05	.00	
	LIB_49	.29	0.04	.00	
	LIB_50	.54	0.04	.00	
	LIB_34	.57	0.04	.00	
	LIB_37	.67	0.04	.00	
	LIB_38	.71	0.03	.00	
LIB_45	.54	0.04	.00		
LIB_46	.49	0.04	.00		
LIB_47	.33	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	了解母語使用者間的交談	LIIB_1	.52	0.05	.00
		LIIB_2	.67	0.04	.00
		LIIB_3	.43	0.05	.00
		LIIB_4	.71	0.03	.00
		LIIB_5	.66	0.04	.00
		LIIB_6	.47	0.04	.00
		LIIB_7	.65	0.04	.00
		LIIB_8	.63	0.04	.00
		LIIB_9	.31	0.05	.00
		LIIB_10	.39	0.05	.00
		LIIB_11	.60	0.04	.00
		LIIB_12	.57	0.04	.00
		LIIB_13	.65	0.04	.00
		LIIB_14	.44	0.05	.00
		LIIB_15	.25	0.05	.00
		LIIB_16	.60	0.04	.00
		LIIB_17	.46	0.04	.00
		LIIB_18	.56	0.04	.00
		LIIB_19	.35	0.05	.00
		LIIB_20	.26	0.05	.00
		LIIB_21	.46	0.05	.00
		LIIB_22	.53	0.04	.00
		LIIB_23	.45	0.05	.00
		LIIB_24	.25	0.05	.00
		LIIB_25	.47	0.05	.00
		LIIB_26	.51	0.04	.00
		LIIB_27	.61	0.04	.00
		LIIB_28	.56	0.04	.00
		LIIB_29	.37	0.05	.00
		LIIB_30	.54	0.04	.00
	聆聽語音媒體與錄音	LIIB_31	.59	0.05	.00
		LIIB_32	.39	0.06	.00
		LIIB_34	.61	0.04	.00
		LIIB_39	.52	0.04	.00
		LIIB_40	.65	0.04	.00
		LIIB_41	.47	0.04	.00
		LIIB_42	.50	0.05	.00
		LIIB_43	.35	0.05	.00
	LIIB_44	.34	0.06	.00	
	作為現場聽眾一員的聽力表現	LIIB_33	.48	0.07	.00
		LIIB_35	.50	0.05	.00
		LIIB_36	.71	0.04	.00
		LIIB_37	.58	0.04	.00
		LIIB_38	.46	0.05	.00
		LIIB_45	.42	0.05	.00
		LIIB_46	.39	0.05	.00
		LIIB_47	.26	0.05	.00
		LIIB_48	.48	0.05	.00
		LIIB_49	.43	0.06	.00
	LIIB_50	.09	0.05	.10	

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	了解母語使用者間的交談	LIC_1	.55	0.06	.00
		LIC_2	.51	0.06	.00
		LIC_3	.57	0.06	.00
		LIC_4	.50	0.06	.00
		LIC_5	.52	0.06	.00
		LIC_6	.54	0.06	.00
		LIC_7	.70	0.04	.00
		LIC_8	.47	0.06	.00
		LIC_9	.31	0.07	.00
		LIC_10	.65	0.05	.00
		LIC_11	.53	0.06	.00
		LIC_12	.52	0.06	.00
		LIC_13	.57	0.05	.00
		LIC_14	.50	0.06	.00
		LIC_15	.54	0.07	.00
		LIC_16	.68	0.05	.00
		LIC_17	.56	0.06	.00
		LIC_18	.51	0.06	.00
		LIC_19	.49	0.06	.00
		LIC_20	.61	0.05	.00
		LIC_21	.69	0.05	.00
		LIC_22	.53	0.05	.00
		LIC_23	.56	0.05	.00
		LIC_24	.52	0.06	.00
		LIC_25	.42	0.06	.00
		LIC_26	.52	0.06	.00
	LIC_27	.55	0.06	.00	
	LIC_28	.62	0.05	.00	
	LIC_29	.52	0.06	.00	
	LIC_30	.43	0.06	.00	
	LIC_31	.38	0.06	.00	
	LIC_32	.26	0.07	.00	
	LIC_33	.61	0.05	.00	
	LIC_34	.56	0.05	.00	
	LIC_35	.25	0.07	.00	
	LIC_36	.48	0.06	.00	
	LIC_40	.62	0.06	.00	
	LIC_41	.55	0.05	.00	
	LIC_42	.66	0.05	.00	
	LIC_43	.59	0.05	.00	
	LIC_44	.38	0.06	.00	
	LIC_45	.33	0.08	.00	
	LIC_46	.43	0.06	.00	
	LIC_37	.50	0.06	.00	
	LIC_38	.63	0.06	.00	
	LIC_39	.70	0.05	.00	
	LIC_47	.62	0.06	.00	
	LIC_48	.55	0.06	.00	
	LIC_49	.36	0.07	.00	
	LIC_50	.65	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	了解母語使用者間的交談	LIIC_1	.44	0.07	.00
		LIIC_2	.61	0.05	.00
		LIIC_3	.65	0.06	.00
		LIIC_4	.41	0.07	.00
		LIIC_5	.47	0.07	.00
		LIIC_6	.61	0.06	.00
		LIIC_7	.58	0.07	.00
		LIIC_8	.62	0.05	.00
		LIIC_9	.68	0.05	.00
		LIIC_10	.67	0.05	.00
		LIIC_11	.55	0.06	.00
		LIIC_12	.49	0.07	.00
		LIIC_16	.54	0.08	.00
		LIIC_17	.42	0.07	.00
		LIIC_18	.59	0.06	.00
		LIIC_27	.42	0.07	.00
		LIIC_28	.50	0.07	.00
		LIIC_29	.52	0.06	.00
		LIIC_30	.45	0.07	.00
		LIIC_13	.51	0.06	.00
	LIIC_14	.48	0.06	.00	
	LIIC_15	.52	0.06	.00	
	LIIC_19	.49	0.07	.00	
	LIIC_20	.64	0.05	.00	
	LIIC_21	.69	0.05	.00	
	LIIC_22	.71	0.05	.00	
	LIIC_23	.46	0.07	.00	
	LIIC_24	.55	0.06	.00	
	LIIC_25	.31	0.07	.00	
	LIIC_26	.62	0.06	.00	
	LIIC_31	.58	0.05	.00	
	LIIC_32	.33	0.07	.00	
	LIIC_33	.38	0.07	.00	
	LIIC_34	.61	0.06	.00	
	LIIC_37	.51	0.08	.00	
	LIIC_38	.48	0.06	.00	
LIIC_39	.65	0.06	.00		
LIIC_40	.73	0.06	.00		
LIIC_41	.69	0.05	.00		
LIIC_42	.75	0.05	.00		
LIIC_43	.57	0.07	.00		
LIIC_44	.57	0.06	.00		
LIIC_45	.65	0.05	.00		
LIIC_46	.66	0.05	.00		
LIIC_47	.59	0.06	.00		
LIIC_48	.57	0.06	.00		
LIIC_49	.55	0.06	.00		
LIIC_50	.48	0.06	.00		
聽宣布及說明指導	LIIC_35	.84	0.07	.00	
	LIIC_36	.75	0.08	.00	

書名：華語文能力測驗技術報告—2013(1)
聽力測驗信效度

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