華語文能力測驗技術報告—2018（1）

聽力測驗、閱讀測驗信效度

國家華語測驗推動工作委員會

目錄

壹、 前言 4
貳、 測驗介紹 5
一、 測驗說明 5
（一） 能力描述 5
（二） 測驗題型 7
（三） 通過門檻 9
二、 等級對照 11
三、 測驗標準化流程 12
參、 測驗效能分析 13
一、 信度分析 13
（一） 內部一致性信度 13
（二） 測驗訊息量 14
二、 效度分析 16
（一） 建構效度 16
（二） 效標關聯效度 20
肆、 結論 22
伍、 文獻 23
陸、 附件 24
附件 1 聽力測驗、閱讀測驗正式考試驗測驗訊息量及測量誤差 24
附件 2 聽力測驗正式考試驗證性因素分析結果 28
附件 3 閱讀測驗正式考試驗證性因素分析結果 32
附件 4 自評問卷 36

表目錄

表 1 聽讀測驗基本能力描述 6

表 2 聽力測驗題型 8

表 3 閱讀測驗題型 8

表 4 聽力測驗、閱讀測驗通過分數 10

表 5 測驗標準化流程說明 12

表 6 聽力測驗與閱讀測驗各等級庫李信度係數 14

表 7 聽力測驗與閱讀測驗各等級門檻測驗訊息量、測量誤差及信度 15

表 8 測驗試題適配分布 17

表 9 聽力測驗、閱讀測驗整體模式適配度指標摘要表 20

表 10 聽力測驗自評問卷各題與測驗總分之相關分析 20

表 11 閱讀測驗自評問卷各題與測驗總分之相關分析 21

圖目錄

圖 1 聽力測驗等級對照圖 11

圖 2 閱讀測驗等級對照圖 11

圖 3 測驗標準化流程 12

圖 4 入門基礎級聽力測驗測驗訊息量 15

圖 5 入門基礎級聽力測驗測量誤差 15

圖 6 入門基礎級聽力測驗三因素模式 18

圖 7 入門基礎級閱讀測驗三因素模式 19

壹、前言

「華語文能力測驗」為一套專為母語非華語者所研發的標準化語言能力測驗，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。華語文能力測驗的測驗類別包括華語文聽力測驗、華語文閱讀測驗、華語文口語測驗、華語文寫作測驗以及兒童華語文能力測驗，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，提供語言學習者能夠衡量其語言能力的國際評量工具。

華語文聽力測驗與閱讀測驗將語言能力分成四等八級，四等分別為準備級、入門基礎級、進階高階級及流利精通級，每一等又再細分為兩級，分別為準備級一級、準備級二級、入門級、基礎級、進階級、高階級、流利級、精通級，共八級。應試者依據自己的學習背景或語言能力選擇合適的等級應考，只要參加一次測驗，即可同時判斷兩等級程度。此測驗架構不僅能區分應試者是否通過測驗，更能進一步區分出通過測驗的應試者群的能力高低。對於應試者及試務工作者來說，更符合經濟效益。

本報告首先簡介華語文聽力測驗與華語文閱讀測驗之能力指標與測驗內容，再針對本年度華語文測驗實施，聽力測驗、閱讀測驗各等級通過門檻與製卷、成績公布之標準化流程進行概述。最後分別闡述正式考試之信度及效度分析結果。

貳、測驗介紹

一、測驗說明

華語文聽力測驗與華語文閱讀測驗正式考試等級分為四等八級，準備級測驗（Band Novice）對應美國外語教學協會指標（ACTFL Proficiency Guidelines）Novice Low 與 Novice Mid；入門基礎級（Band A）、進階高階級（Band B）與流利精通級（Band C）則分別對應歐洲共同語文參考架構（CEFR）A1（Breakthrough）、A2（Waystage）、B1（Threshold）、B2（Vantage）、C1（Effective Operational Proficiency）與 C2（Mastery）（Council of Europe,2001）。華語文能力測驗等級為準備級（Band Novice），包括準備級一級（N1）、準備級二級（N2）；入門基礎級（Band A），包括入門級（Level 1）、基礎級（Level 2）；進階高階級（Band B），包括進階級（Level 3）、高階級（Level 4）；以及流利精通級（Band C），包括流利級（Level 5）、精通級（Level 6）。

以下將就聽力測驗與閱讀測驗之通過等級能力描述、測驗題型與題數、以及通過門檻四方面進行介紹。

（一）能力描述

聽力測驗和閱讀測驗各等級通過者所需具備的基本聽力能力和閱讀能力如下表所示。其中，準備級著重在「語言基本單位的識別能力」；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」；流利精通級則著重在「語言使用的廣度與精熟度」。各等級所應具備的語言能力說明如表 1：

表 1 聽讀測驗基本能力描述

通過等級	聽力測驗能力描述	閱讀測驗能力描述
準備級一級	在發音清晰、語速緩慢、可重聽，且有語境支持的情況下，能辨別出日常生活中與個人切身相關的高頻詞彙。	在可重複閱讀且有語境支持的情況下，能辨別出日常生活中與個人切身相關的高頻詞彙。
準備級二級	在發音清晰、語速緩慢、可重聽，且有語境支持的情況下，能聽懂日常生活中所使用的高頻詞彙或短語。	在可重複閱讀且有語境支持的情況下，能理解日常生活中所使用的高頻詞彙或短語。
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。	在有視覺協助及可重複閱讀的情況下，能掌握簡單的短語並能大致理解句子內容。
基礎級	當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。	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	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。	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	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。	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	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。	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	能聽懂語言不規範或帶有生僻專業術語的演講、廣播和報告等口頭連貫敘述的內容。	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

聽力測驗和閱讀測驗為接收型之語言測驗，兩項測驗均以各等級語言學習者之能力表現及所預設之測知目標為參考依據，為各等級設計題型。

準備級語言使用者無論在聽力或閱讀的能力上都僅有辨識與抓取高頻詞彙的能力，需倚賴大量視覺輔助或語言之外的協助，才能了解接收到的訊息。因此，此階段測驗題型大部分採用圖像式測驗，或提供圖文相輔的視覺輔助，透過圖片呈現任務。

入門基礎級語言使用者從「詞彙」進入到「句子」與「有組織的篇章材料」。考量此階段的語言使用者無論在聽力或閱讀的能力上仍需倚賴視覺輔助或語言之外的協助，才能了解接收到的訊息。因此，此階段測驗題型部分採用圖像式測驗，或提供圖文相輔的視覺輔助，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，在此範疇下的，語言使用者應能掌握以標準語表達的聽力材料、談話內容，或具有一定篇幅的閱讀文本，能理解這些訊息的大意和重要細節。因此，此等級題型設計中，聽力測驗除較簡短的對話外，也加入較長篇幅的對話及段落，閱讀測驗則以真實材料和短文為題幹的題型為主，選項皆以文字呈現，取代圖片設計。

流利精通級的語言使用者具備高度穩定的語言能力，不管文本結構或關聯性是否清楚明確，都能理解並掌握要點與重要細節。因此針對此階段的應試者，不論是聽力的文本或閱讀的文本均篇幅較長、體裁更多元，內容取材範圍也更深入至專業領域的範疇。

綜上所述，華語文聽力測驗準備級分為三個部分：第一部分、第二部分、第三部分；入門基礎級分為四個部分：看圖回答、問答理解、對話理解與對話，前三個部分為三選一的單選題，第四部分對話為四選一的單選題；進階高階級和流利精通級則分為二個部分：對話與段落。

而華語文閱讀測驗準備級分為兩個部分：第一部分、第二部分；入門基礎級

分為五個部分：單句理解、看圖釋義、選詞填空、完成段落與閱讀理解；進階高階級與流利精通級則分為二個部分：選詞填空與閱讀理解。各等級題型分布請見表 2、表 3：

表 2 聽力測驗題型

測驗等級	題型
準備級	第一部分
	第二部分
	第三部分
入門基礎級	看圖回答
	問答理解
	對話理解
	對話
進階高階級	對話
	段落
流利精通級	對話
	段落

表 3 閱讀測驗題型

測驗等級	題型
準備級	第一部分
	第二部分
入門基礎級	單句理解
	看圖釋義
	選詞填空
	完成段落
	閱讀理解
進階高階級	選詞填空
	閱讀理解
流利精通級	選詞填空
	閱讀理解

（三）通過門檻

華語文聽力測驗與華語文閱讀測驗共透過六次標準設定（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（藍珮君等人，2013；國家華語測驗推動工作委員會，2015）。所有參與標準設定之成員均由華語文教學及語言學領域的專家學者組成，且每一次標準設定程序皆依循相同標準化流程執行。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，驗證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A1至C2六個等級的有效性。

準備級華語文聽力與閱讀測驗透過標準設定（standard setting）程序，於2017年5月辦理兩場研究會議，各別邀請華語文、語言學領域專家及資深華語教師，採用書籤法（Lewis, Miztel, & Green, 1996）制定出準備級一級與準備級二級之通過門檻。

依據標準設定結果，應試者的測驗結果與等級通過判定依據以量尺分數（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，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。四等測驗之量尺分數滿分皆為80分，而各等級通過分數如表4 所示：

表 4 聽力測驗、閱讀測驗通過分數

測驗等級	通過等級	聽力測驗通過分數	閱讀測驗通過分數
準備級	準備級一級	38	39
	準備級二級	59	61
入門基礎級	入門級	41	42
	基礎級	60	60
進階高階級	進階級	46	48
	高階級	61	64
流利精通級	流利級	50	52
	精通級	61	69

二、等級對照

圖 1、圖 2 為華語文聽力測驗和閱讀測驗分級方式及與歐洲共同語文參考架構（CEFR）、美國外語教學協會指標（ACTFL Proficiency Guidelines）之對應關係[1]。

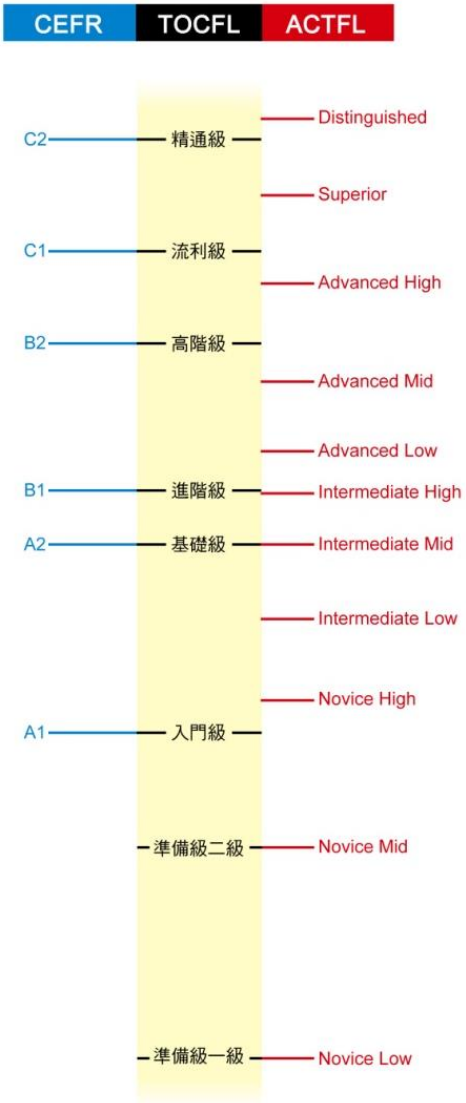

圖 1 聽力測驗等級對照圖

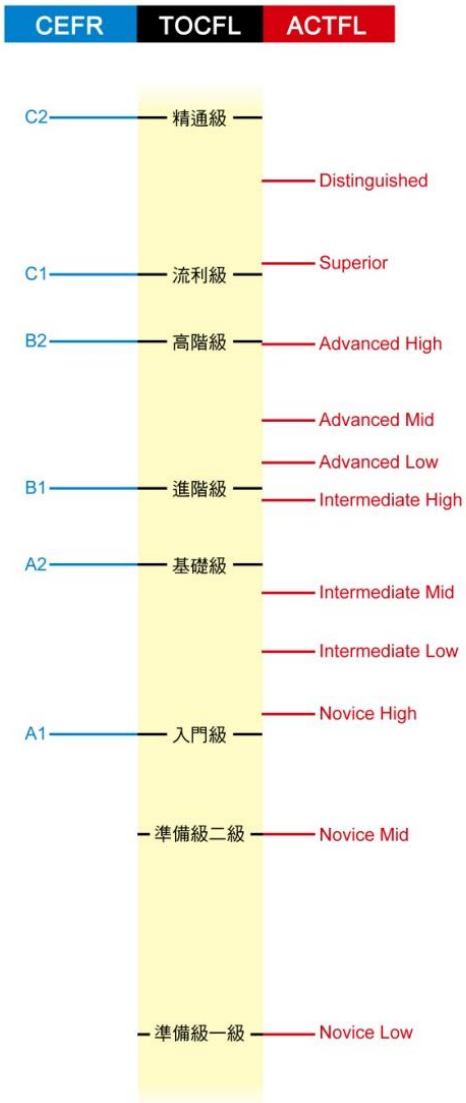

圖 2 閱讀測驗等級對照圖

[1]資料來源為本會於 2012 至 2017 年進行之標準設定系列研究成果。

三、測驗標準化流程

華語文能力測驗標準化流程如圖 3 所示,「測驗標準化」各步驟說明如表 5。共包含：試題收集、修審、題庫輸入、組預試卷、試題分析、組合正式卷、成績檢核與口寫評分、考後結果分析等八個步驟，加上對外舉辦的預試考試、正式考試與考後成績公佈，完成整套「測驗標準化流程」。

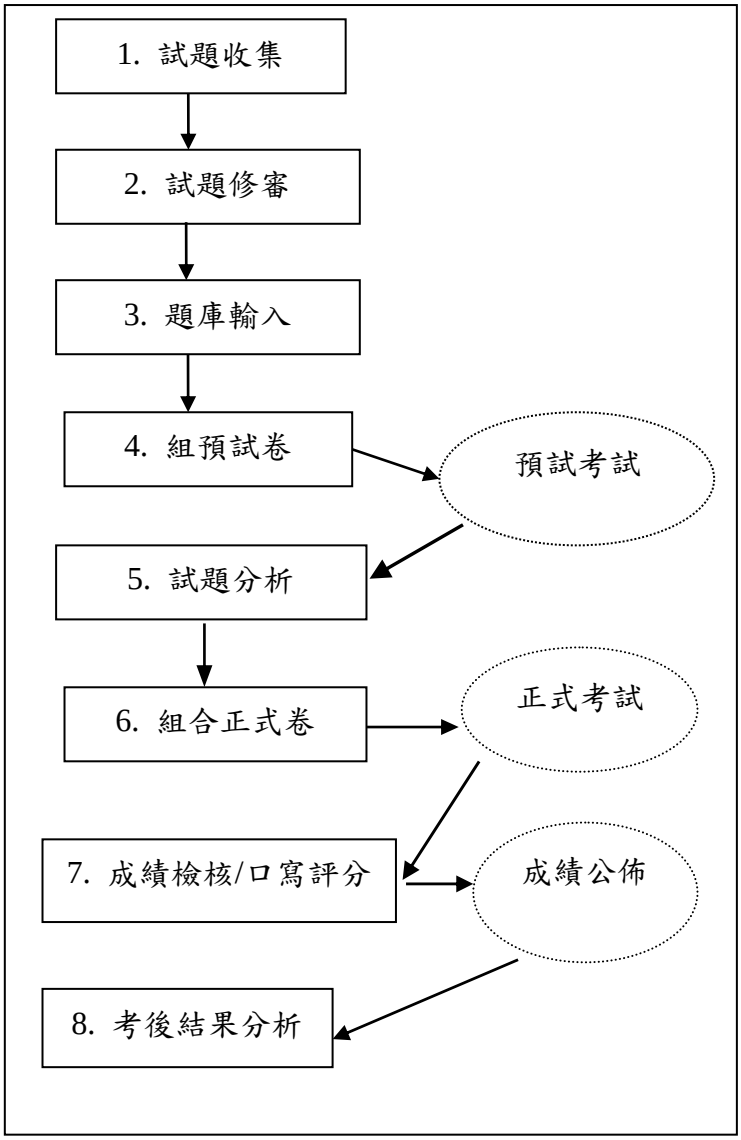

圖 3 測驗標準化流程

表 5 測驗標準化流程說明

項目	說明
1. 試題收集	辦理命題研習;與命題教師進行試題設計溝通 。
2. 試題修審	組修審稿、繪圖、三階段審查（會內初審、專家審查、會內複審）聯繫審查專家;依據審查意見修稿。
3. 題庫輸入	音檔、影片檔錄製與檢核;試題校對。
4. 組預試卷	紙筆測驗:製作繁體版與繁簡版題本；多國語題本。 電腦測驗:題目由題庫產出電子卷包，放上考試系統進行繁、簡版題目檢核。
5. 試題分析/統計數值輸入	成績檢核;試題統計分析;試題討論(確定保留的題目將標記為入庫題，並輸入統計分析數值，以利後續組正式卷)。
6. 組合正式卷	紙筆測驗：同「組預試卷」。 電腦測驗：測驗包含 CAT、CBT、快篩，檢核同「組預試卷」。
7. 成績檢核	各測驗成績報表檢核：聽讀口寫、兒童測驗與統計分析分別針對考後各項成績報表進行檢核,確認成績無誤，對外公告成績。
8. 考後結果分析	考後各類報表彙整。 進行下列項目分析:正式考試測驗成績分布及通過比例分析;測驗信效度分析 。

參、測驗效能分析

華語文聽力測驗與閱讀測驗的組卷方式是依據試題反應理論（Item Response Theory；以下簡稱 IRT）而來的，IRT 之一重要假設為單向度假設（uni-dimensionality）；亦即測驗中所有題目皆測量相同潛在特質（或能力），當考生並非靠所欲測知的單一特質（或能力）作答時，單向度 IRT 分析所獲得的試題參數及考生能力估計值將是偏誤的。上述測驗藉由 Winsteps 3.68.2 版[2]進行單向度 IRT 分析測驗相關資料，分析結果可區分為信度分析與效度分析。本年度以五月國內正式考試資料進行分析。

一、信度分析

以下從內部一致性及測驗訊息量（test information；TI*(θ)*）兩個面向評估聽力測驗以及閱讀測驗的信度。

（一）內部一致性信度

測驗的信度是指測驗結果的一致性和穩定性，亦即測驗題目是否一致測量到相同的潛在特質。內部一致性信度是常見的測驗信度評估指標之一，它是指同一份測驗中所有試題的相關程度有多高。如果所有試題皆測量到相同的目標特質，那麼每道試題測量結果的相關性應該很高（陳柏熹，2011）[3]，因此，內部一致性指標可以顯示內容取樣及內容異質性的誤差，以反映內容取樣對於測驗分數的影響。各級測驗之內部一致性信度將以庫李 20 號公式（Kuder-Richardson）為指標。

由表 6 顯示，聽力測驗三等級信度係數介於.81 至.87 之間，閱讀測驗信度係數介於.83 至.92 之間，顯示本年度度正式測驗之庫李信度大致良好。

[2] Linacre, J.M. (2009). Winsteps® (Version 3.68.2) [Computer Software]. Beaverton, Oregon: Winsteps.com.
[3] 陳柏熹(2011)。心理與教育測驗：測驗編製理論與實務。臺北：精策教育。

表 6 聽力測驗與閱讀測驗各等級庫李信度係數

測驗類型	測驗等級	題數	信度係數
聽力	入門基礎級	45	.81
	進階高階級	45	.82
	流利精通級	45	.87
閱讀	入門基礎級	44	.90
	進階高階級	44	.83
	流利精通級	45	.92

（二）測驗訊息量

測驗訊息量表示測量精準度，訊息量越高，表示試題對於考生能力的測量精準度越高。較詳細的理論及公式請見華測會出版之《華語文能力測驗技術報告 2013-1 聽力測驗信效度》[4]第四章第一節。

圖 4、圖 5 為入門基礎級聽力測驗測驗訊息量及測量誤差，其中，測量誤差均轉換為量尺分數。因篇幅有限，其他測驗類型及等級之測驗訊息量與測量誤差圖請見附件 1，由附件 1 圖 1-1 至圖 1-10 可知，各測驗等級正式考試在通過門檻皆具有較高的測驗訊息量與較低的測量誤差，並在通過門檻均具有大致良好的信度係數，顯示測驗信度良好。

聽力與閱讀測驗各等級正式考試通過門檻的測驗訊息量、測量誤差及轉換後的信度係數如表 7 所示，各等級各項測驗結果相仿，以入門基礎級測驗為例，聽力測驗入門級與基礎級門檻的測驗訊息量分別為 8.31 與 5.93，測量誤差分別為 3.47 與 4.11，信度係數分別為.87 與.81；閱讀測驗正式考試入門級與基礎級門檻的測驗訊息量分別為 7.61 與 6.24，測量誤差分別為 3.62 與 4.00，信度係數分別為.94 與.92。

[4] 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 2013-1 聽力測驗信效度。新北市：國家華語測驗推動工作委員會。

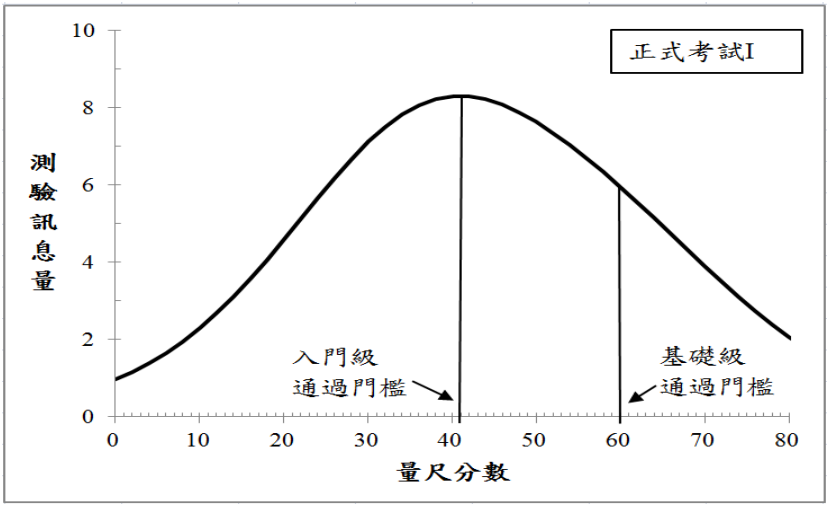

圖 4 入門基礎級聽力測驗測驗訊息量

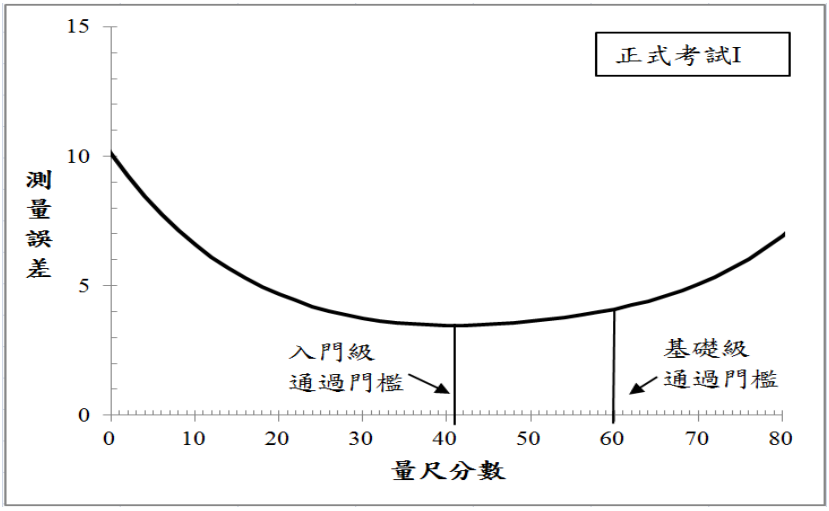

圖 5 入門基礎級聽力測驗測量誤差

表 7 聽力測驗與閱讀測驗各等級門檻測驗訊息量、測量誤差及信度

測驗類型	等級	測驗訊息量	測量誤差	轉換後信度係數
聽力	入門級/基礎級	8.31 / 5.93	3.47 / 4.11	.87 / .81
	進階級/高階級	8.01 / 6.97	3.53 / 3.79	.84 / .81
	流利級/精通級	9.31 / 7.09	3.28 / 3.76	.89 / .85
閱讀	入門級/基礎級	7.61 / 6.24	3.62 / 4.00	.94 / .92
	進階級/高階級	7.50 / 7.01	3.65 / 3.78	.86 / .85
	流利級/精通級	9.86 / 5.39	3.18 / 4.31	.94 / .89

二、效度分析

為評估試題所測量的能力是否與測驗發展所訂定的架構內容相吻合，且是否測量到所欲測量的能力，聽力測驗與閱讀測驗的效度分析藉由試題分析、因素分析，評估測驗的建構效度。聽力、閱讀測驗另透過考生自評能力與測驗表現之間的關係來評估效標關聯效度。

（一）建構效度

由 IRT 試題分析與驗證性因素分析結果可知，華語文聽力測驗與華語文閱讀測驗各項測驗各等級皆具有一定之建構效度。

1. IRT 試題分析

由於預試時考生答題動機可能不如正式考試，加上正式考試考生整體能力未必與試題難度相當，考生整體能力有時高於試題難度，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，以 Winsteps 3.68.2 版進行單向度 IRT 分析，採用 MNSQ 介於 0.6 到 1.4 以及 ZSTD 介於-3.0 到 3.0 的標準評估試題是否與單向度 IRT 模式適配，亦即 MNSQ 以及 ZSTD 超出設定範圍者，表示題目不符合單向度 IRT 模式。各項測驗各等級試題適配分布如表 8 所示，聽力測驗三個等級正式考試試題適配率依序為 80%、96%、98%；閱讀測驗則分別為89%、84%、96%。各次考試大多數的試題適配率皆超過 80%，表示大部分試題皆測量到相同潛在特質，具有建構效度。 未來將持續追蹤不適配試題於正式卷的分析結果，不適配次數頻繁的試題後續將修改後重新預試。

表 8 測驗試題適配分布

測驗類型	測驗等級	總題數/向度	適配題數	適配率
聽力	入門基礎級	45	36	80%
	進階高階級	45	43	96%
	流利精通級	45	44	98%
閱讀	入門基礎級	44	39	89%
	進階高階級	44	37	84%
	流利精通級	45	43	96%

2. 驗證性因素分析

此節使用 Mplus 7.0 版[5]進行驗證性因素分析，使用之估計方法與相關評估指標請參考華測會出版之《華語文能力測驗技術報告 2013-1 聽力測驗信效度》第四章第二節。聽力測驗與閱讀測驗各等級的因素結構乃參考 CEFR 聽力和閱讀活動分項能力描述與測驗題型、內容設計可測得之聽力理解與閱讀理解能力而擬定。

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，兩項測驗之各等級具有一定的建構效度，各道試題分別可測得不同的分項能力，而各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和閱讀理解能力。由於篇幅有限，在此僅呈現入門基礎級聽力測驗與讀測驗的結果，其餘等級的因素結構圖及各等級各題因素負荷量等參數請見附件 2、附件 3。

入門基礎級聽力測驗正式考試驗證性因素分析結果如圖 6 所示，各試題的所屬能力分別為：

1.「在有視覺輔助的前提下，能聽懂簡單的問題」（LAI_01-10）

2.「在有視覺輔助的前提下，能聽懂日常生活中簡短的對話或宣布，並掌握關鍵訊息」（LAI_11-36）

3.「能理解日常生活對話中具體、清楚的訊息」（LAI_37-45）。

在基本適配指標部分，三因素模式驗證性因素分析結果顯示，試題因素負荷量介於.08 至.82 之間，平均值為.50，78%的試題因素負荷量達.40 以上；除 LAI_44 與 LAI_45 外，其餘各題因素負荷量統計考驗均達顯著水準（$p<.05$）；潛在變項

[5] Muthén, L.K. and Muthén, B.O. (2012). Mplus® (Version 7.0) [Computer Software]. Los Angeles, CA: Muthén & Muthén.

間的相關係數分別為.88、.54 及.84。

圖 7 為入門基礎級閱讀測驗因素模式圖，各試題的所屬能力分別為：

1.「能讀懂非常基本、常用的詞彙和簡單的句子」（RAI_01-18, RAI_20-22, RAI_26-27, RAI_29,RAI_31-35）

2.「能在常見的廣告、宣傳品、菜單、時刻表中找到自己所需的特定信息」（RAI_19, RAI_23-25, RAI_28, RAI_30）

3.「能讀懂大部分簡單文本（如記敘文、私人信件）中的要點」（RAI_36-44）。

三因素模式驗證性因素分析結果顯示，試題因素負荷量介於.14 至.86 之間，平均值為.63，91%的試題因素負荷量達.40 以上；各題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分別為.98、.82 及.88。

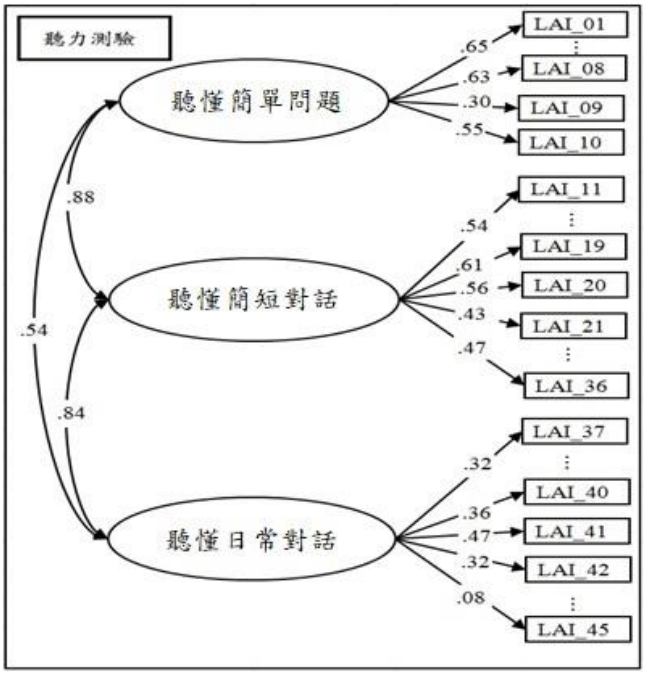

圖 6 入門基礎級聽力測驗三因素模式

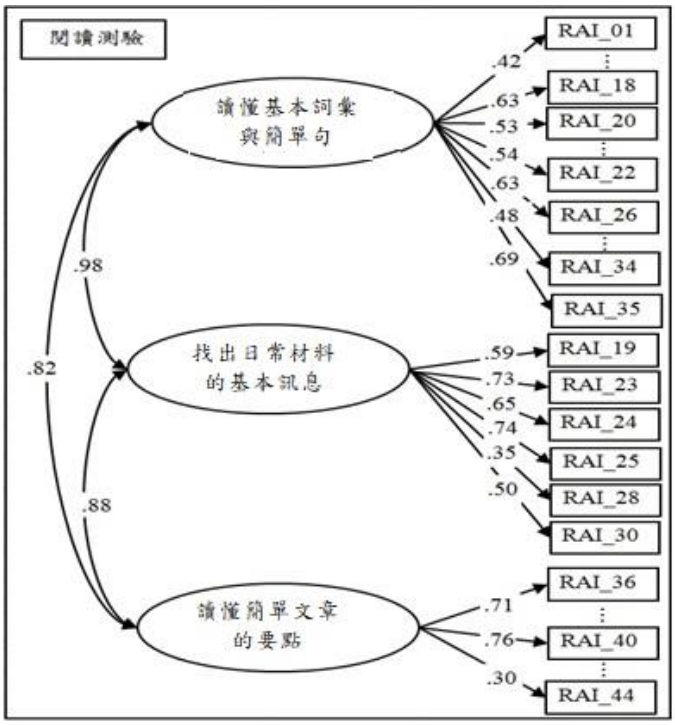

圖 7 入門基礎級閱讀測驗三因素模式

透過整體模式適配度分析，可評鑑整個模式與觀察資料的適合程度。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（邱正皓，2011）[6]。由於聽力、閱讀測驗在進行驗證性因素分析時，自由度大多接近 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 9 可知，各項測驗各等級的卡方自由度比均小於 3，表示模式適配度良好。絕對適配度評估的分析結果，聽讀測驗平方概似平方誤根係數（root mean square error of approximation，以下簡稱 RMSEA）均小於.08，表示大多數模式皆符合絕對適配度指標。增值適配度評估部分，各類型各等級測驗的比較適配指標（comparative-fit index，簡稱 CFI）和非規範適配指標（non-normed fit index，簡稱 NNFI）均接近或大於.90，顯示模式均符合增值適配度指標。

[6] 邱皓政(2011)。結構方程模式：LISREL / SIMPLIS 原理與應用。臺北：雙葉書廊。

表 9 聽力測驗、閱讀測驗整體模式適配度指標摘要表

測驗類型	檢驗模式	χ^2/df	RMSEA	CFI	NNFI
聽力	入門基礎級三因素模式	1.17	0.018	.94	.94
	進階高階級二因素模式	1.31	0.019	.95	.95
	流利精通級二因素模式	1.08	0.018	.97	.97
閱讀	入門基礎級三因素模式	1.20	0.019	.98	.97
	進階高階級三因素模式	1.83	0.031	.88	.88
	流利精通級二因素模式	1.11	0.021	.98	.98

（二）效標關聯效度

為了瞭解考生對於自己華語能力表現評估與實際測驗表現之間的關聯性，在聽力測驗與閱讀測驗進階高階級與流利精通級正式考試[7]結束後，請考生各填答一份自評問卷，就問卷與測驗成績進行相關分析。分析結果顯示，各項測驗考生自評表現與測驗結果之間大體上皆有正相關存在，顯示具有效標關聯效度。

聽力測驗在進階高階級方面，正式考試考生自評結果與測驗總分的積差相關係數為.403（p<.01）；流利精通級方面，正式考試考生自評結果與測驗總分的積差相關係數為.499（p<.01），表示自評聽力能力越佳者，其聽力測驗分數越高，具有效標關聯效度。此外，自評問卷各題回答結果與測驗總分的斯皮爾曼等級相關分析結果如表 10 所示。進階高階級各題與測驗總分的相關係數介於.319 至.375 之間（p<.01）；流利精通級各題與測驗總分的相關係數介於.455 至.550 間（p<.01），表示在這些題目中回答可以做到頻率越高的考生，其測驗總分也越高。詳細問卷內容請見附件 4。

表 10 聽力測驗自評問卷各題與測驗總分之相關分析

進階高階級	Q1	Q2	Q3	Q4	Q5
（N=472）	.331**	.375**	.357**	.357**	.319**
流利精通級	Q1	Q2	Q3	Q4	
（N=206）	.550**	.455**	.459**	.506**	

**p<.01

[7] 入門基礎級因需提供多國語問卷，系統尚無法支援，故先針對進階高階級、流利精通級考生進行調查。

閱讀測驗在進階高階級方面，正式考試考生自評結果與測驗總分的積差相關係數為.368（p<.01）；流利精通級方面，正式考試考生自評結果與測驗總分的積差相關係數為.576（p<.01），表示自評閱讀能力越佳者，其閱讀測驗分數越高，具有效標關聯效度。將自評問卷各題回答結果與測驗總分進行等級相關分析，結果如表 11 所示。進階高階級各題與測驗總分的相關係數介於.301 至.363 之間（p<.01）；流利精通級各題與測驗總分的相關係數介於.475 至.606 間（p<.01），表示在這些題目中回答可以做到頻率越高的考生，其測驗總分也越高。詳細問卷內容請見附件 4。

表 11 閱讀測驗自評問卷各題與測驗總分之相關分析

進階高階級（N=472）	Q1	Q2	Q3	Q4	Q5	
	.363**	.322**	.314**	.303**	.301**	
流利精通級（N=206）	Q1	Q2	Q3	Q4	Q5	Q6
	.523**	.593**	.565**	.475**	.520**	.606**

**p<.01

肆、結論

本文主要針對華語文聽力測驗和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、施測和成績公布之標準化流程。同時檢視本年度全國性正式考試信度與效度之測驗效能評估。

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「測驗內容效度」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，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，最後，輔以考生考後自評問卷與測驗結果的相關分析做為「效標關聯效度」證據。經由信度及效度分析結果，顯示各等級聽讀測驗結果已相當可靠且皆具有建構效度，發揮了測驗效能。

伍、文獻

陳柏熹(2011)。心理與教育測驗：測驗編製理論與實務。臺北：精策教育。

邱皓政(2011)。結構方程模式：LISREL / SIMPLIS 原理與應用。臺北：雙葉書廊。

國家華語測驗推動工作委員會(2014)。華語文能力測驗技術報告 2014 華語文聽力測驗技術報告。新北市：國家華語測驗推動工作委員會。

國家華語測驗推動工作委員會(2014)。華語文能力測驗技術報告 2014 華語文閱讀測驗技術報告。新北市：國家華語測驗推動工作委員會。

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.

Lewis, D. M., Mitzel, H.C., & Green, D. R. (1996). S*tandard setting: A bookmark approach*. Paper presented at the Council of Chief State School Officers National Conference on Large Scale Assessment, Boulder, CO.

Linacre, J.M. (2009). Winsteps® (Version 3.68.2) [Computer Software]. Beaverton, Oregon: Winsteps.com.

Impara, J. C., & Plake, B. S. (1997). Standard-setting: An alternative approach. *Journal of Educational Measurement*, *34*, 353–366.

Muthén, L.K. and Muthén, B.O. (2012). Mplus® (Version 7.0) [Computer Software]. Los Angeles, CA: Muthén & Muthén.

陸、附件

附件 1 聽力測驗、閱讀測驗正式考試驗測驗訊息量及測量誤差

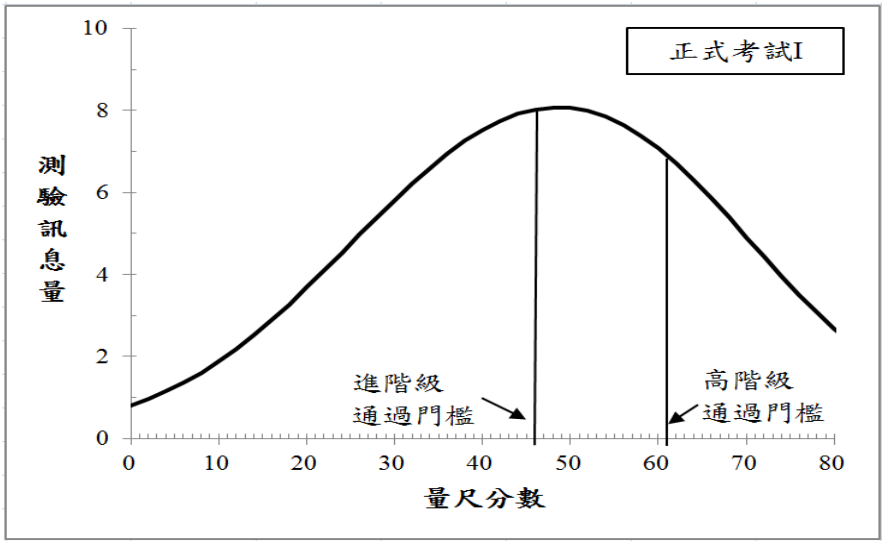

圖 1-1 進階高階級聽力測驗測驗訊息量

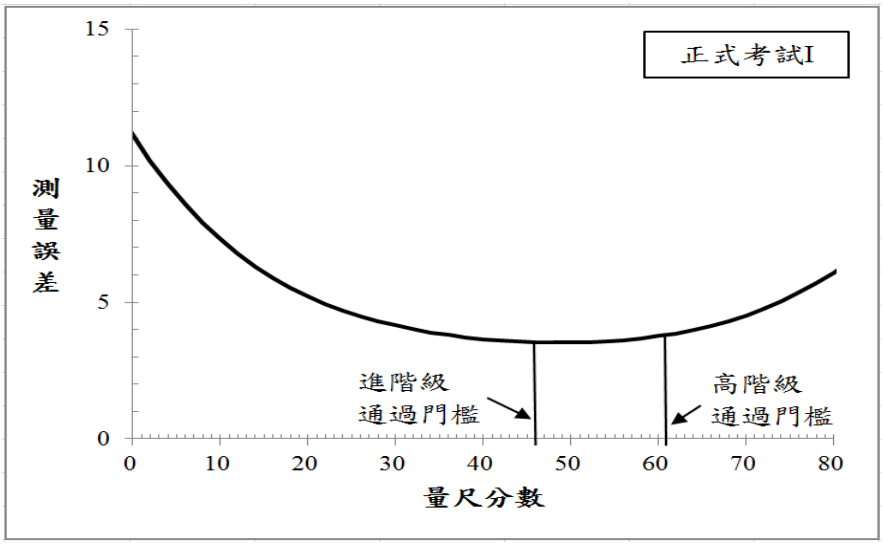

圖 1-2 進階高階級聽力測驗測量誤差

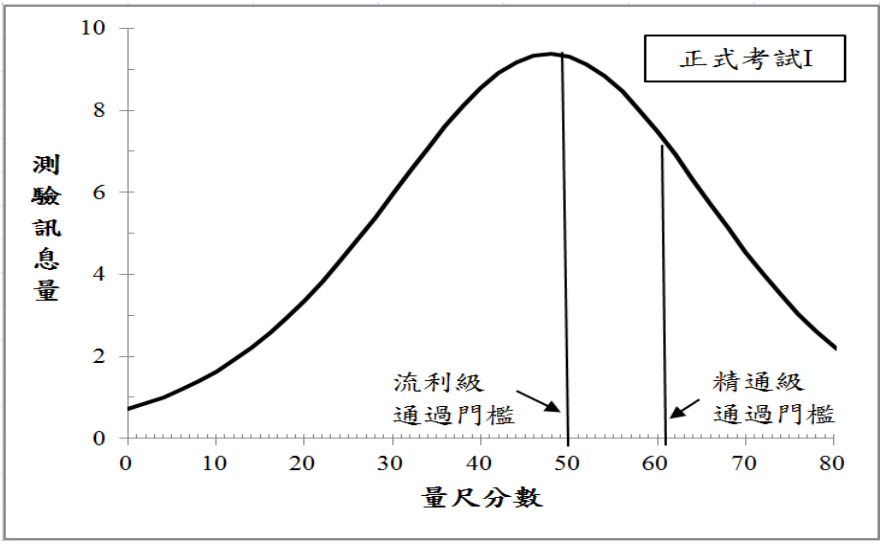

圖 1-3 流利精通級聽力測驗測驗訊息量

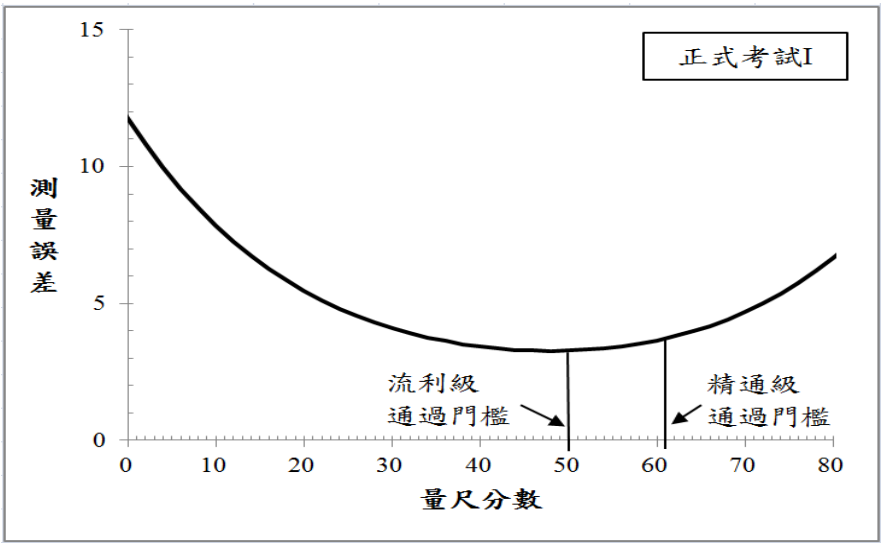

圖 1-4 流利精通級聽力測驗測量誤差

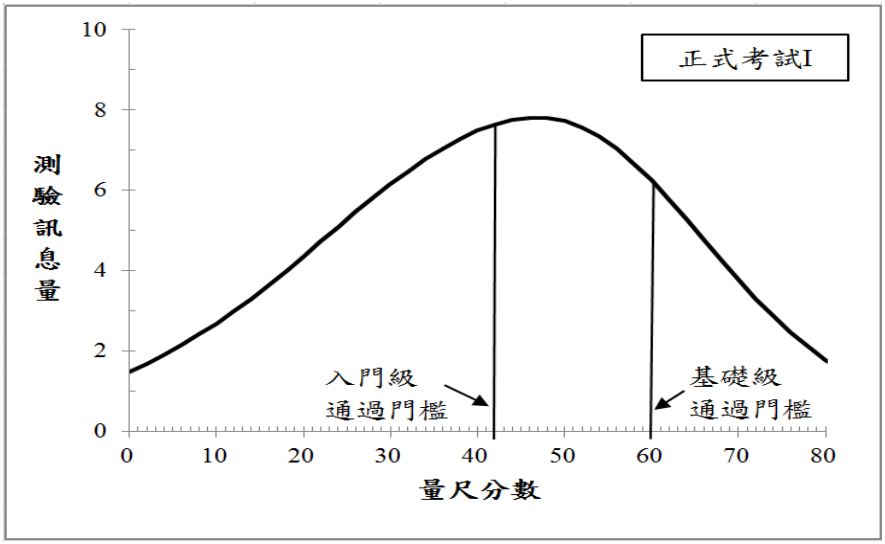

圖 1-5 入門基礎級閱讀測驗測驗訊息量

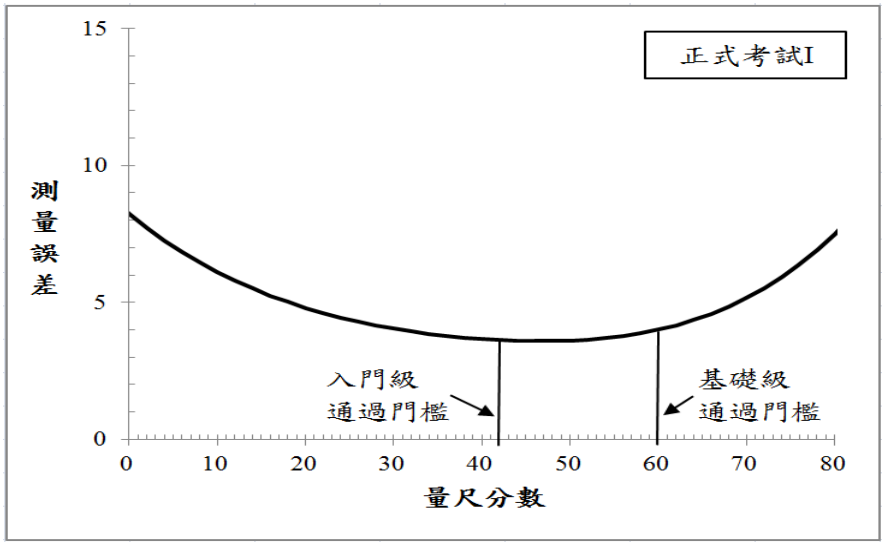

圖 1-6 入門基礎級閱讀測驗測量誤差

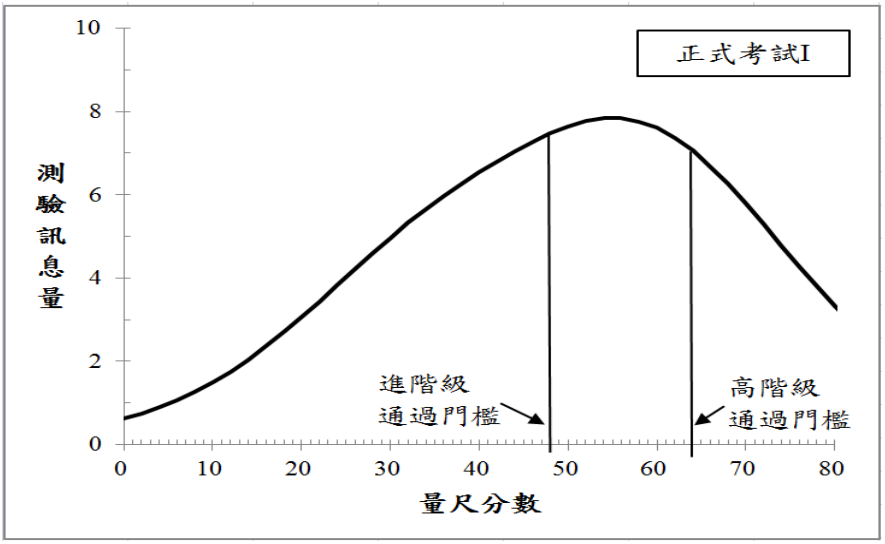

圖 1-7 進階高階級閱讀測驗測驗訊息量

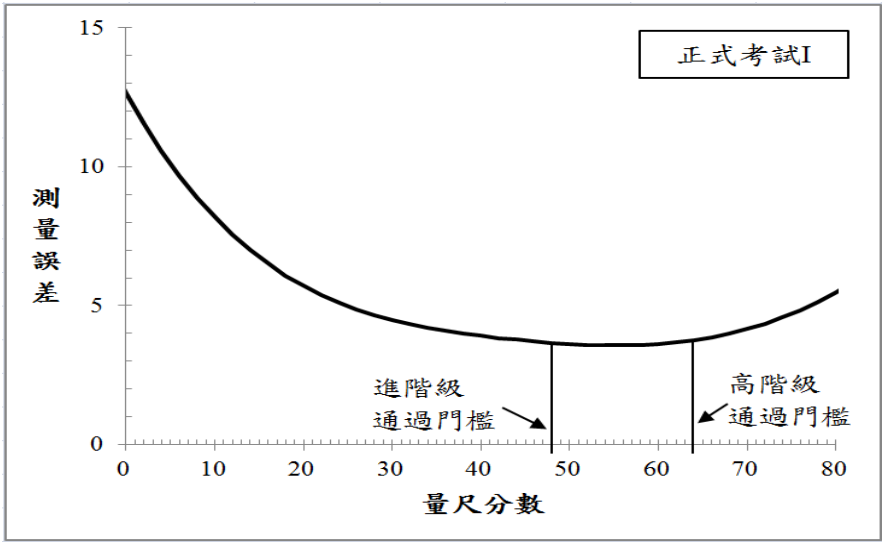

圖 1-8 進階高階級閱讀測驗測量誤差

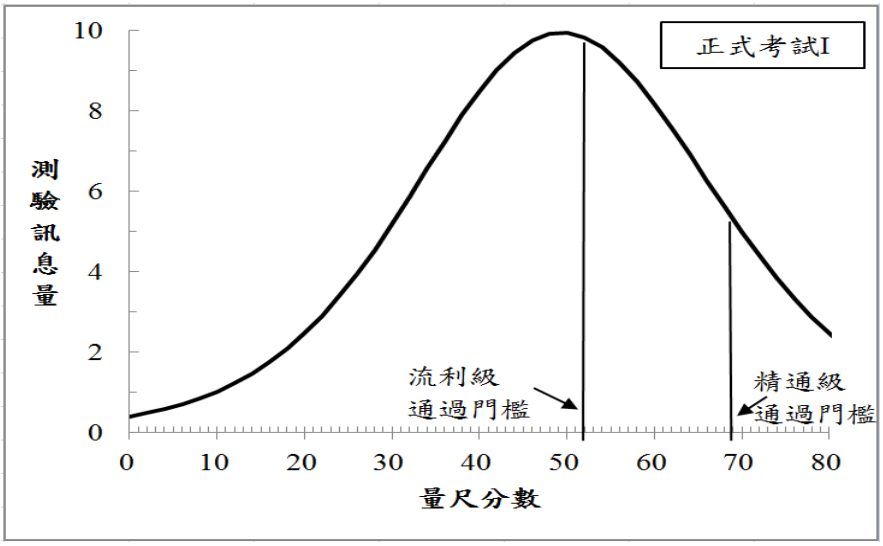

圖 1-9 流利精通級閱讀測驗測驗訊息量

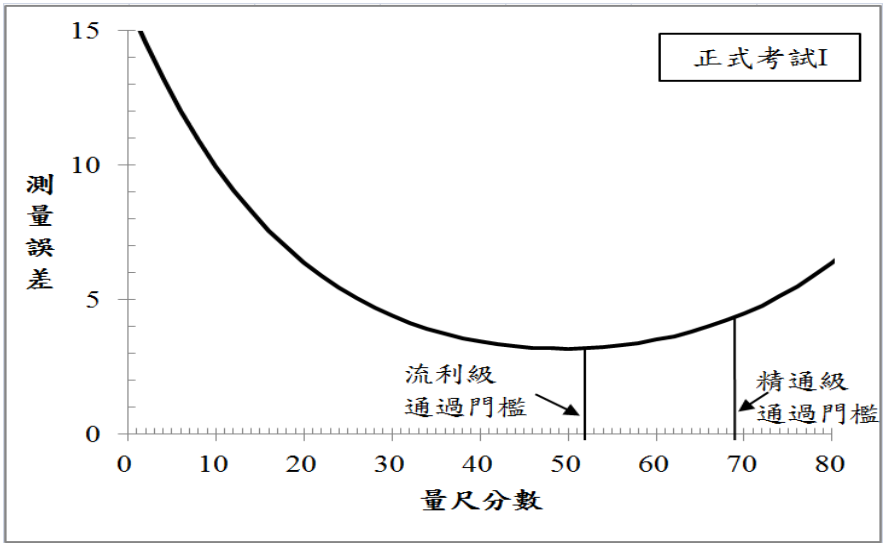

圖 1-10 流利精通級閱讀測驗測量誤差

附件 2 聽力測驗正式考試驗證性因素分析結果

表 2-1 入門基礎級測驗各題因素負荷量及殘差變異量摘要表

分項能力	試題編號	因素負荷量	標準誤	P 值
在有視覺輔助的前提下，能聽懂簡單的問題	LAI_01	0.65	0.062	0.000
	LAI_02	0.46	0.121	0.000
	LAI_03	0.38	0.112	0.001
	LAI_04	0.82	0.059	0.000
	LAI_05	0.54	0.052	0.000
	LAI_06	0.78	0.054	0.000
	LAI_07	0.50	0.096	0.000
	LAI_08	0.63	0.053	0.000
	LAI_09	0.30	0.069	0.000
	LAI_10	0.55	0.063	0.000
在有視覺輔助的前提下，能聽懂日常生活中簡短的對話或宣布，並掌握關鍵訊息	LAI_11	0.54	0.059	0.000
	LAI_12	0.49	0.067	0.000
	LAI_13	0.58	0.066	0.000
	LAI_14	0.47	0.072	0.000
	LAI_15	0.65	0.053	0.000
	LAI_16	0.52	0.063	0.000
	LAI_17	0.38	0.060	0.000
	LAI_18	0.48	0.058	0.000
	LAI_19	0.61	0.056	0.000
	LAI_20	0.56	0.056	0.000
	LAI_21	0.43	0.057	0.000
	LAI_22	0.42	0.052	0.000
	LAI_23	0.26	0.060	0.000
	LAI_24	0.61	0.052	0.000
	LAI_25	0.50	0.084	0.000
	LAI_26	0.73	0.044	0.000
	LAI_27	0.70	0.053	0.000
	LAI_28	0.35	0.067	0.000
	LAI_29	0.58	0.051	0.000
	LAI_30	0.60	0.064	0.000
	LAI_31	0.62	0.060	0.000
	LAI_32	0.55	0.050	0.000
	LAI_33	0.59	0.045	0.000
	LAI_34	0.42	0.057	0.000
	LAI_35	0.79	0.039	0.000
	LAI_36	0.47	0.055	0.000
能理解日常生活對話中具體、清楚的訊息	LAI_37	0.32	0.069	0.000
	LAI_38	0.45	0.066	0.000
	LAI_39	0.47	0.065	0.000
	LAI_40	0.36	0.068	0.000
	LAI_41	0.47	0.074	0.000
	LAI_42	0.32	0.076	0.000
	LAI_43	0.44	0.070	0.000
	LAI_44	0.12	0.073	0.108
	LAI_45	0.08	0.072	0.288

表 2-2 進階高階級測驗各題因素負荷量及殘差變異量摘要表

分項能力	試題編號	因素負荷量	標準誤	P 值
能理解對話內容及延伸討論的大意；能掌握對話中重要的訊息與細節；並能辨識出說話者的態度、立場與觀點	LBI_01	0.62	0.054	0.000
	LBI_02	0.75	0.052	0.000
	LBI_03	0.62	0.056	0.000
	LBI_04	0.55	0.057	0.000
	LBI_05	0.70	0.037	0.000
	LBI_06	0.57	0.051	0.000
	LBI_07	0.57	0.046	0.000
	LBI_08	0.55	0.056	0.000
	LBI_09	0.65	0.043	0.000
	LBI_10	0.61	0.045	0.000
	LBI_11	0.56	0.040	0.000
	LBI_12	0.44	0.047	0.000
	LBI_13	0.50	0.045	0.000
	LBI_14	0.46	0.042	0.000
	LBI_15	0.47	0.041	0.000
	LBI_16	0.22	0.049	0.000
	LBI_17	0.33	0.047	0.000
	LBI_18	0.04	0.054	0.460
	LBI_19	0.43	0.052	0.000
	LBI_20	0.28	0.047	0.000
	LBI_21	0.33	0.049	0.000
	LBI_22	0.27	0.047	0.000
	LBI_23	0.51	0.045	0.000
	LBI_24	0.54	0.040	0.000
	LBI_25	0.50	0.045	0.000
在聆聽廣播、新聞報導或宣布時，能理解談話內容的重點和大意；對結構清楚的演說，能掌握內容重點和大意	LBI_26	0.59	0.042	0.000
	LBI_27	0.52	0.043	0.000
	LBI_28	0.57	0.040	0.000
	LBI_29	0.41	0.047	0.000
	LBI_30	0.43	0.047	0.000
	LBI_31	0.48	0.043	0.000
	LBI_32	0.65	0.036	0.000
	LBI_33	0.61	0.037	0.000
	LBI_34	0.36	0.046	0.000
	LBI_35	0.16	0.048	0.001
	LBI_36	0.49	0.042	0.000
	LBI_37	0.41	0.043	0.000
	LBI_38	0.40	0.048	0.000
	LBI_39	0.60	0.037	0.000
	LBI_40	0.45	0.043	0.000
	LBI_41	0.52	0.039	0.000
	LBI_42	0.18	0.066	0.006
	LBI_43	0.06	0.053	0.284
	LBI_44	0.02	0.058	0.719
	LBI_45	0.31	0.053	0.000

表 2-3 流利精通級測驗各題因素負荷量及殘差變異量摘要表

分項能力	試題編號	因素負荷量	標準誤	P 值
能理解並掌握對話中長且複雜的內容與重要細節，包含說話者的立場、觀點、看法，以及隱含的態度	LCI_01	0.49	0.069	0.000
	LCI_02	0.65	0.068	0.000
	LCI_03	0.54	0.067	0.000
	LCI_04	0.66	0.058	0.000
	LCI_05	0.66	0.057	0.000
	LCI_06	0.65	0.064	0.000
	LCI_07	0.67	0.062	0.000
	LCI_08	0.61	0.059	0.000
	LCI_09	0.71	0.052	0.000
	LCI_10	0.38	0.074	0.000
	LCI_11	0.49	0.070	0.000
	LCI_12	-0.21	0.087	0.015
	LCI_13	0.39	0.080	0.000
	LCI_14	0.72	0.051	0.000
	LCI_15	0.33	0.076	0.000
	LCI_16	0.67	0.069	0.000
	LCI_17	0.68	0.055	0.000
	LCI_18	0.39	0.075	0.000
	LCI_19	0.55	0.076	0.000
	LCI_20	0.50	0.070	0.000
	LCI_21	0.58	0.074	0.000
	LCI_22	0.61	0.058	0.000
	LCI_23	0.49	0.074	0.000
	LCI_24	0.56	0.066	0.000
	LCI_25	0.34	0.076	0.000
能理解語言形式複雜或帶有專業術語的各種口頭連貫敘述的內容	LCI_26	0.50	0.100	0.000
	LCI_27	0.55	0.066	0.000
	LCI_28	0.23	0.086	0.008
	LCI_29	0.50	0.072	0.000
	LCI_30	0.73	0.060	0.000
	LCI_31	0.64	0.067	0.000
	LCI_32	0.20	0.083	0.017
	LCI_33	0.00	0.085	0.998
	LCI_34	0.22	0.082	0.008
	LCI_35	0.70	0.054	0.000
	LCI_36	0.62	0.061	0.000
	LCI_37	0.74	0.055	0.000
	LCI_38	0.80	0.047	0.000
	LCI_39	0.55	0.071	0.000
	LCI_40	0.52	0.068	0.000
	LCI_41	0.37	0.077	0.000
	LCI_42	0.55	0.069	0.000
	LCI_43	0.26	0.099	0.008
	LCI_44	0.67	0.060	0.000
	LCI_45	0.73	0.068	0.000
	LCI_46	0.09	0.082	0.254

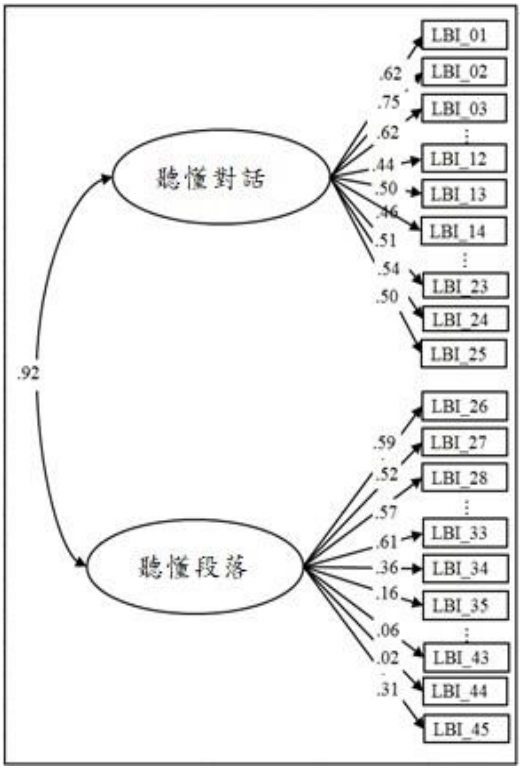

圖 2-1 進階高階級聽力測驗二因素模式

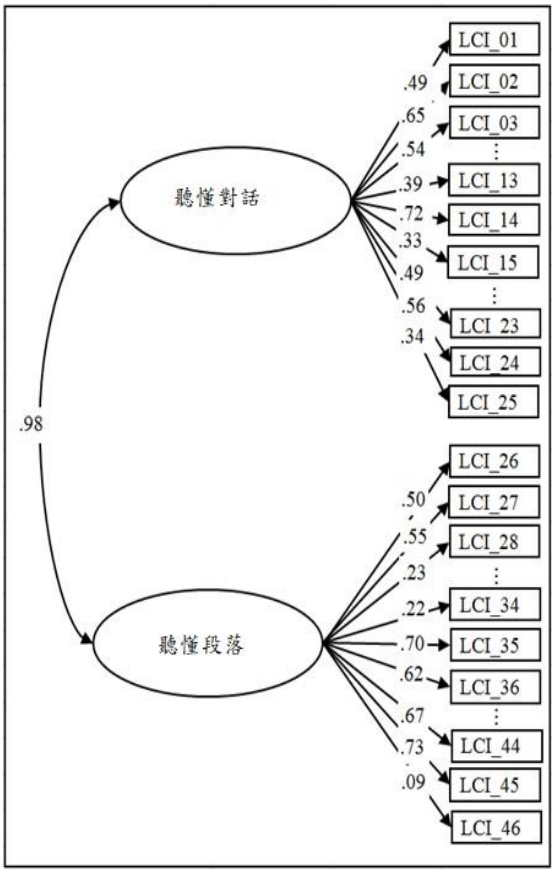

圖 2-2 流利精通級聽力測驗二因素模式

附件 3 閱讀測驗正式考試驗證性因素分析結果

表 3-1 入門基礎級測驗各題因素負荷量及殘差變異量摘要表

分項能力	試題編號	因素負荷量	標準誤	P 值
能讀懂非常基本、常用的詞彙和簡單的句子	RAI_01	0.42	0.113	0.000
	RAI_02	0.61	0.076	0.000
	RAI_03	0.86	0.069	0.000
	RAI_04	0.62	0.072	0.000
	RAI_05	0.64	0.065	0.000
	RAI_06	0.83	0.048	0.000
	RAI_07	0.61	0.068	0.000
	RAI_08	0.70	0.064	0.000
	RAI_09	0.62	0.088	0.000
	RAI_10	0.58	0.072	0.000
	RAI_11	0.64	0.076	0.000
	RAI_12	0.63	0.046	0.000
	RAI_13	0.14	0.060	0.017
	RAI_14	0.72	0.039	0.000
	RAI_15	0.50	0.052	0.000
	RAI_16	0.84	0.038	0.000
	RAI_17	0.79	0.046	0.000
	RAI_18	0.63	0.056	0.000
	RAI_20	0.53	0.061	0.000
	RAI_21	0.69	0.050	0.000
	RAI_22	0.54	0.054	0.000
	RAI_26	0.63	0.050	0.000
	RAI_27	0.69	0.042	0.000
	RAI_29	0.78	0.034	0.000
	RAI_31	0.34	0.066	0.000
	RAI_32	0.58	0.049	0.000
	RAI_33	0.73	0.038	0.000
	RAI_34	0.48	0.053	0.000
	RAI_35	0.69	0.045	0.000
能在常見的廣告、宣傳品、菜單、時刻表中找到自己所需的特定信息	RAI_19	0.59	0.053	0.000
	RAI_23	0.73	0.043	0.000
	RAI_24	0.65	0.049	0.000
	RAI_25	0.74	0.041	0.000
	RAI_28	0.35	0.055	0.000
	RAI_30	0.50	0.051	0.000
能讀懂大部分簡單文本(如記敘文、私人信件)中的要點	RAI_36	0.71	0.045	0.000
	RAI_37	0.72	0.043	0.000
	RAI_38	0.63	0.051	0.000
	RAI_39	0.69	0.044	0.000
	RAI_40	0.76	0.040	0.000
	RAI_41	0.71	0.047	0.000
	RAI_42	0.71	0.048	0.000
	RAI_43	0.72	0.048	0.000
	RAI_44	0.30	0.065	0.000

表 3-2 進階高階級測驗各題因素負荷量及殘差變異量摘要表

分項能力	試題編號	因素負荷量	標準誤	P 值
能讀懂以日常用語描寫的事件、情感和願望	RBI_13	0.67	0.037	0.000
	RBI_14	0.43	0.042	0.000
	RBI_15	0.35	0.052	0.000
	RBI_16	0.36	0.049	0.000
	RBI_17	0.61	0.061	0.000
	RBI_18	0.44	0.044	0.000
	RBI_19	0.55	0.041	0.000
	RBI_20	0.46	0.042	0.000
	RBI_21	0.49	0.043	0.000
	RBI_22	0.60	0.039	0.000
	RBI_23	0.45	0.051	0.000
	RBI_24	0.30	0.056	0.000
	RBI_31	0.39	0.052	0.000
	RBI_32	0.29	0.050	0.000
能從明白易懂的文章中找出所需要的訊息	RBI_01	0.31	0.046	0.000
	RBI_02	0.46	0.042	0.000
	RBI_03	0.41	0.047	0.000
	RBI_04	0.25	0.046	0.000
	RBI_05	0.35	0.050	0.000
	RBI_06	0.68	0.030	0.000
	RBI_07	0.43	0.044	0.000
	RBI_08	0.57	0.037	0.000
	RBI_09	0.52	0.045	0.000
	RBI_10	0.25	0.048	0.000
	RBI_11	0.64	0.034	0.000
	RBI_12	0.58	0.041	0.000
	RBI_25	0.50	0.040	0.000
	RBI_26	0.51	0.039	0.000
	RBI_27	0.56	0.040	0.000
	RBI_28	0.49	0.044	0.000
	RBI_29	0.64	0.035	0.000
	RBI_30	0.26	0.047	0.000
	RBI_33	0.43	0.050	0.000
	RBI_34	0.14	0.052	0.006
	RBI_35	0.36	0.044	0.000
	RBI_36	0.41	0.046	0.000
	RBI_37	0.69	0.034	0.000
能讀懂帶有作者個人立場與見解，且與當前議題相關的文章和報導	RBI_38	0.65	0.041	0.000
	RBI_39	0.64	0.052	0.000
	RBI_40	0.49	0.049	0.000
	RBI_41	0.44	0.048	0.000
	RBI_42	0.59	0.047	0.000
	RBI_43	0.55	0.053	0.000
	RBI_44	0.51	0.049	0.000

表 3-3 流利精通級測驗各題因素負荷量及殘差變異量摘要表

分項能力	試題編號	因素負荷量	標準誤	P 值
不管文本的主題是否熟悉，能讀懂一篇長且複雜的文章及技術說明	RCI_18	0.74	0.050	0.000
	RCI_19	0.71	0.052	0.000
	RCI_20	0.60	0.062	0.000
	RCI_21	0.79	0.044	0.000
	RCI_22	0.32	0.077	0.000
	RCI_23	0.51	0.070	0.000
	RCI_24	-0.12	0.086	0.164
	RCI_25	0.67	0.058	0.000
	RCI_26	0.47	0.070	0.000
	RCI_27	0.61	0.061	0.000
	RCI_28	0.61	0.060	0.000
	RCI_35	0.46	0.070	0.000
	RCI_40	0.69	0.058	0.000
	RCI_41	0.67	0.055	0.000
	RCI_42	0.64	0.061	0.000
	RCI_43	0.70	0.051	0.000
	RCI_44	0.66	0.056	0.000
	RCI_45	0.69	0.052	0.000
能讀懂各類文章，包括帶有許多成語、俗語或內容抽象的文本，並能讀懂文中的言外之意	RCI_01	0.82	0.050	0.000
	RCI_02	0.65	0.059	0.000
	RCI_03	0.66	0.054	0.000
	RCI_04	0.55	0.066	0.000
	RCI_05	0.73	0.046	0.000
	RCI_06	0.78	0.041	0.000
	RCI_07	0.65	0.054	0.000
	RCI_08	0.74	0.052	0.000
	RCI_09	0.64	0.057	0.000
	RCI_10	0.71	0.053	0.000
	RCI_11	0.63	0.062	0.000
	RCI_12	0.57	0.068	0.000
	RCI_13	0.66	0.056	0.000
	RCI_14	0.76	0.044	0.000
	RCI_15	0.56	0.064	0.000
	RCI_16	0.01	0.082	0.934
	RCI_17	0.63	0.058	0.000
	RCI_29	0.57	0.064	0.000
	RCI_30	0.78	0.042	0.000
	RCI_31	0.63	0.059	0.000
	RCI_32	0.38	0.075	0.000
	RCI_33	0.61	0.058	0.000
	RCI_34	0.76	0.046	0.000
	RCI_36	0.36	0.089	0.000
	RCI_37	0.41	0.072	0.000
	RCI_38	0.43	0.071	0.000
	RCI_39	0.40	0.076	0.000

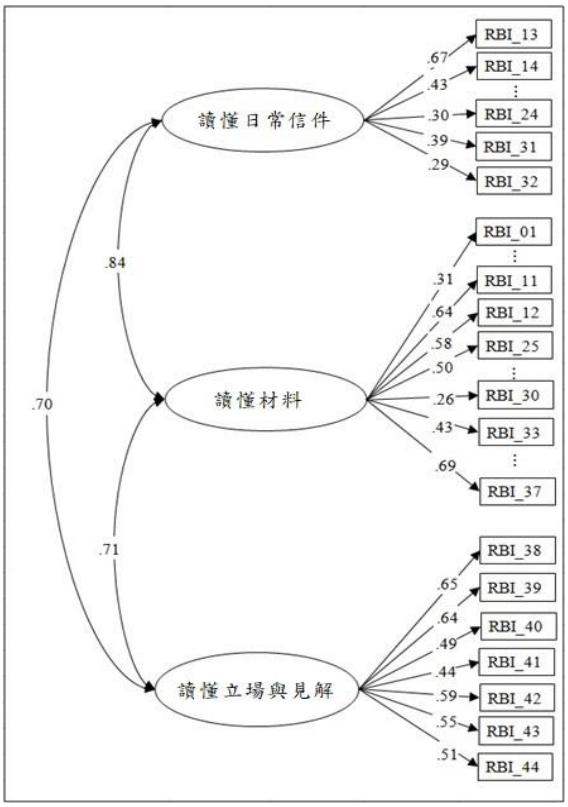

圖 3-1 進階高階級閱讀測驗三因素模式

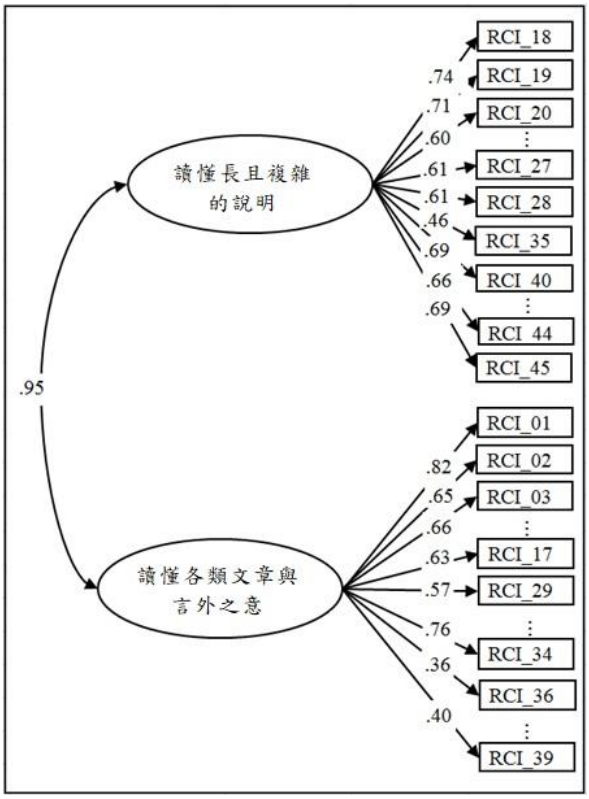

圖 3-2 流利精通級閱讀測驗二因素模式

附件 4 自評問卷

附件 4-1 聽力測驗自評問卷

- 進階高階級自評問卷

一、我能聽懂以簡單連貫的語言所進行的日常生活對話。

□從不　□很少　□有時　□常常　□總是

二、我能正確分辨出說話者的觀點、態度及談話內容。

□從不　□很少　□有時　□常常　□總是

三、我能聽懂電視新聞、記錄片及訪談節目大部分的內容。

□從不　□很少　□有時　□常常　□總是

四、我能聽懂常見話題和短篇敘述的主要內容。

□從不　□很少　□有時　□常常　□總是

五、我能聽懂報導中的大意和細節。

□從不　□很少　□有時　□常常　□總是

- 流利精通級自評問卷

一、我能聽懂交談者之間全部的對話。

□從不　□很少　□有時　□常常　□總是

二、即使話題抽象、複雜或不常見，我能輕易地聽懂討論或辯論中的複雜內容。

□從不　□很少　□有時　□常常　□總是

三、我能察覺微妙的細節，包括說話者隱含的態度。

□從不　□很少　□有時　□常常　□總是

四、我能正確地分辨說話者的情緒、語氣等等。

□從不　□很少　□有時　□常常　□總是

附件 4-2 閱讀測驗自評問卷

- 進階高階級自評問卷

一、我能讀懂以日常用語描寫的情感和願望。

□從不　□很少　□有時　□常常　□總是

二、我能整合來自一或多段文字中的不同資訊。

□從不　□很少　□有時　□常常　□總是

三、我能從一個長篇文章（如信件、傳單、官方文件和報紙）的不同部分中找出重要的訊息。

□從不　□很少　□有時　□常常　□總是

四、我能讀懂帶有作者個人立場與見解，且與當前議題相關的文章和報導。

□從不　□很少　□有時　□常常　□總是

五、我能讀懂文章的細節。

□從不　□很少　□有時　□常常　□總是

- 流利精通級自評問卷

一、我能讀懂與我專業領域無關的專業文章和長篇技術說明。

□從不　□很少　□有時　□常常　□總是

二、不管文本的主題是否熟悉，如果能重新閱讀困難的部分，我能讀懂一篇長且複雜的文章。

□從不　□很少　□有時　□常常　□總是

三、我能看懂不同類型的文學作品（如詩、散文、戲劇）。

□從不　□很少　□有時　□常常　□總是

四、我能讀懂涉及不熟悉領域的科學文章中的細節。

□從不　□很少　□有時　□常常　□總是

五、我能讀懂含有許多成語和俗語的文本。

□從不　□很少　□有時　□常常　□總是

六、我能讀懂言外之意和文字遊戲，如文本中具有諷刺意味的文字。

□從不　□很少　□有時　□常常　□總是