

華語文能力測驗技術報告—2024（2）

口語測驗、寫作測驗信效度

目錄

壹、前言	5
貳、測驗介紹	6
一、測驗說明	6
(一) 能力描述	7
(二) 測驗題型	9
(三) 通過門檻與評分規準	11
二、測驗標準化流程	13
參、測驗效能分析	15
一、評分者信度	15
(一) 評分者間信度 (嚴格度分析)	16
(二) 評分者間信度 (斯皮爾曼等級相關)	20
(三) 評分者內信度	22
二、效度分析	25
(一) 建構效度	25
(二) 效標關聯效度	28
(三) 程序性效度	31
肆、結論	35
伍、參考文獻	36
陸、附件	37
附件 1 口語、寫作測驗正式考試驗證性因素分析結果	37
附件 2 口語測驗考生自評問卷	39
附件 3 寫作測驗流利精通級問卷	48
附件 4 評分會議評分人員調查問卷	49

表目錄

表 1	口語和寫作測驗基本能力描述	8
表 2	口語測驗題型	10
表 3	寫作測驗題型	11
表 4	口語測驗通過分數	12
表 5	寫作測驗通過分數	12
表 6	測驗標準化流程說明	14
表 7	評分流程說明	14
表 8	入門基礎級口語測驗評分者嚴格度（3 月）	17
表 9	入門基礎級口語測驗評分者嚴格度（5 月）	18
表 10	進階高階級口語測驗評分者嚴格度	18
表 11	流利精通級口語測驗評分者嚴格度	18
表 12	入門基礎級寫作測驗圖片描述題型評分者嚴格度	19
表 13	入門基礎級寫作測驗書信寫作題型評分者嚴格度	19
表 14	進階高階級寫作測驗書信寫作題型評分者嚴格度	20
表 15	進階高階級寫作測驗觀點論述題型評分者嚴格度	20
表 16	流利精通級寫作測驗摘要寫作題型評分者嚴格度	20
表 17	流利精通級寫作測驗觀點論述題型評分者嚴格度	20
表 18	口語測驗評分者間斯皮爾曼等級相關	21
表 19	寫作測驗評分者間斯皮爾曼等級相關	22
表 20	口語測驗、寫作測驗試題適配分布	26
表 21	口語測驗、寫作測驗整體模式適配度指標摘要表	28
表 22	口語測驗自評問卷各題與測驗總分之相關分析	30
表 23	寫作測驗自評問卷各題與測驗總分之相關分析	31
表 24	口語測驗評分會議流程	32
表 25	寫作測驗評分作業流程	32
表 26	入門基礎級口語測驗評分人員問卷調查結果	33

表 27 入門基礎級寫作測驗評分人員問卷調查結果	34
--------------------------------	----

圖目錄

圖 1 測驗標準化流程	13
圖 2 入門基礎級口語測驗單因素模式（3 月）	27
圖 3 進階高階級寫作測驗單因素模式	27

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。

華語文口語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。華語文寫作測驗考量華語文字結構特性、與語言使用者可達成的任務，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，每一等又再細分為兩級，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來說，更符合經濟效益。

「華語文口語測驗」之命題力求內容之普遍性、真實性，符合一般之交際情境。測驗採電腦化施測，試題透過螢幕和耳機播放，受測者藉由麥克風錄下回答內容並將其回傳至電腦系統。「華語文寫作測驗」則著重於考察受測者能否在特定語境下，藉由書面表達，有效地傳遞訊息；施測形式採電腦化測驗，試題透過螢幕呈現，受測者以鍵盤輸入文字進行寫作。

本報告首先簡介華語文口語測驗與華語文寫作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，口語測驗、寫作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

華語文口語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）。測驗等級為準備級（Band Novice），包括準備級一級（Novice 1）、準備級二級（Novice 2）；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）。

華語文寫作測驗正式考試等級分為三等六級，入門基礎級（Band A）、進階高階級（Band B）與流利精通級（Band C）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）。測驗等級為入門基礎級，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）。

以下將就口語測驗與寫作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

口語測驗與寫作測驗各等級通過者所需具備的基本能力如下表 1 所示。口語測驗準備級著重在「語言基本單位的識別能力」；入門基礎級著重在「針對日常生活中熟悉及例行性的事務進行簡單的回答或描述」、進階高階級著重在「針對自身經驗或感興趣的話題做出有條理的描述或提出看法」、流利精通級則著重在「針對自身領域外抽象、複雜、且具爭議性的主題，發表清楚、有組織的言談」。

寫作測驗入門基礎級著重在「能運用短語或句子表達日常生活中的人、事、物或個人需求」、進階高階級著重在「能撰寫信件或文章描述自身經驗或表達觀點」、流利精通級則著重在「能針對複雜的主題能撰寫文章論述自身的觀點，或重新組織資訊凸顯文本重點」。各等級所應具備的語言能力說明如表 1：

表 1 口語和寫作測驗基本能力描述

通過等級	口語測驗能力描述	寫作測驗能力描述
準備級 一級	在有準備時間的情況下，能說出少數日常生活中與個人切身相關的高頻詞彙。	無可用描述指標。
準備級 二級	在有準備時間的情況下，能使用高頻詞彙，回答日常生活中極為簡單的提問。	無可用描述指標。
入門級	<ol style="list-style-type: none"> 1.能簡短地回答與個人生活密切相關的問題。例如：住在哪裡、認識什麼人、擁有的事物等。 2.能使用熟悉的日常用語與詞彙簡單地描述人物、地點及物品。 	<ol style="list-style-type: none"> 1.能以短語和簡單、不連貫的句子描述與日常生活相關的人物、地點、物品及活動。 2.能以短語和簡單、不連貫的句子書寫與日常生活相關、簡短且非常簡單的訊息，提供資訊或詢問。
基礎級	<ol style="list-style-type: none"> 1.能使用簡單的短語或句子敘述個人背景、日常生活中熟悉的事物及每日例行性事務。 2.能簡單地描述短片的内容。 	<ol style="list-style-type: none"> 1.能以簡單連接詞銜接數個句子，如：而且、可是、因為……。 2.能書寫文意連貫的數個句子簡單地描述與日常生活相關的事件、活動和個人經驗。 3.能書寫文意連貫的較短的信件來交換訊息、回應日常生活相關的問題或表達立即的需求。
進階級	<ol style="list-style-type: none"> 1.能直接且連貫地描述個人相關經驗及感覺、夢想、希望、真實或想像事件。 2.能有次序地說明計畫或事件；能提出簡短的理由支持自己的看法。 	<ol style="list-style-type: none"> 1.能書寫淺白、連貫的文本。 2.能書寫信件詢問資訊、提出要求、給予確認，並提供有限的相關細節，或較詳細地描述經驗、情感、事件。 3.對一般性社會議題，能書寫簡短的文章，敘述事實資訊及行為的原因，能簡單說明自己的立場與觀點。
高階級	<ol style="list-style-type: none"> 1.能清楚、仔細地描述感興趣的話題、經驗或事件。 2.對於一般性議題或有爭議的內容，能提出個人見解、並有組織地詳細說明理由。 3.能發展清晰的論點，舉出相關的例子延伸並支持自己的論點。 	<ol style="list-style-type: none"> 1.能書寫清楚、詳細的文本。 2.能以適切地格式與語體書寫信件，傳達不同程度的情緒，強調事件和經驗對個人的意義，並對通信者傳達的消息或觀點給予評論。 3.對一般性社會議題，能以適切的例證闡述個人支持或反對某特定觀點的理由。

(續)

通過等級	口語測驗能力描述	寫作測驗能力描述
流利級	在正式場合中，針對複雜、抽象、不熟悉的話題：	1.能大致撰寫貼近原意的摘要，對各種主題的長篇文本資料，能大致良好地重新組織，並能大致使用多樣句型與常用書面語。通篇脈絡大致清楚且文句通暢。
	1.能清晰、仔細地描述細節，整合次要主題並做出適當的結論。 2.能流利、自如、適當且具說服力地回應相反論證的內容。 3.能使用其他說明、理由及相關例子，延伸並支持自己的論點。	2.能撰寫充實的長篇文章或報告，對各種議題能具邏輯且清楚地予以評析、提出解決方案與作出結論，以表達個人觀點，並能運用多樣句型與常用書面語。通篇脈絡大致清楚且文句通暢。
精通級	在正式場合中，針對複雜、抽象、不熟悉的話題：	1.能撰寫表達完整原意且突顯重點的摘要，對各種主題的長篇文本資料，能適切地重新組織，並能靈活地運用複雜多樣的句型與豐富的高程度書面語。通篇脈絡清晰且文句簡練。
	1.能發表清楚、流暢、結構完整且具邏輯性的談話，並能幫助聽者掌握重要的部分。 2.能遊刃有餘地以清楚、具說服力的論證維持立場。 3.能彈性地調整說話方式以符合聽者需求。	2.能撰寫充實詳盡的長篇文章或報告，對各種議題能以縝密的邏輯結構予以評析與批判、提出完善的解決方案與作出適當的結論，以表達個人觀點，並能靈活地運用複雜多樣的句型與高程度書面語。通篇脈絡清晰且文句豐富暢達。

(二) 測驗題型

1. 口語測驗

口語測驗為產出型之語言測驗，依照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設計題型。

準備級針對「詞彙」、「短句」的能力，著重提供與個人高度相關的訊息。據此規劃了回答簡單問題。入門基礎級針對「單句層次描述」和「段落層次描述」的能力，著重於提供與個人生活密切相關的信息、簡單回答與工作、閒暇及日常例行生活有關的問題，並描述自身經驗以及對事物的喜好，規劃了回答問題類和描述類兩大題型。進階高階級則針對「做出連貫描述」和「陳述個人看法、論點」的能力規劃了描述類和意見陳述類的兩類題型。而流利精通級則旨在觀察應試者「摘要能力」、「回應能力」以及「論述能力」三大面向的口語能力，並據此設計了五項題型。口語測驗各等級所測題型之題數與各題型的準備時間與回答時間，

請見下表 2：

表 2 口語測驗題型

等級	題型	題數	準備時間	回答時間
準備級	第一部分	10	無	10秒
	第二部分	6	10秒	10秒
入門基礎級	熱身題	2	無	15秒
	回答問題	2	無	30秒
	經驗描述	2	50秒	1分鐘
	影片描述	1	50秒	1分鐘
進階高階級	熱身題	2	無	40秒
	經驗描述	2	50秒	1分40秒
	圖片描述	1	50秒	1分40秒
	陳述意見	3	1分鐘	2分鐘
流利精通級	陳述意見	1	1分鐘	2分鐘
	角色扮演	1	5分鐘	3分鐘
	觀點回應	2	2分鐘	3分鐘
	文章摘要	1	10分鐘	3分鐘
	觀點論述	2	2分鐘	3分鐘

2.寫作測驗

寫作測驗為產出型之語言測驗，依照各等級語言學習者之能力表現及預設之測知目標為參考依據設計題型。

入門基礎級測驗希望測得應試者能「撰寫與日常生活相關人、事、物的簡單描述」，以及能「書寫與個人生活相關的簡短信件」的寫作能力，因此設計了圖片描述與書信寫作兩種題型；進階高階級測驗希望測得應試者能「描述詳細事件、情感、經驗或對個人意義的長篇信件」，以及「針對一般性社會議題能論述特定觀點」的寫作能力，因此設計了書信寫作與觀點論述的兩種題型；流利精通級則希望測得應試者「針對複雜資訊能重新組織並凸顯重點的短篇摘要」，以及「針對各種議題能詳盡闡述自身觀點的長篇文章」的寫作能力，因此設計了摘要寫作與觀點論述兩種題型。寫作測驗各等級所測題型之題數與各題型的準備時間與回

答時間，請見下表 3：

表 3 寫作測驗題型

等級	題型	題數	字數	作答時間
入門基礎級	圖片描述	1	80-100字	20分鐘
	書信寫作	1	80-100字	20分鐘
進階高階級	書信寫作	1	450-550字	60分鐘
	觀點論述	1	450-550字	60分鐘
流利精通級	摘要寫作	1	200-300字	50分鐘
	觀點論述	1	800-1000字	120分鐘

(三) 通過門檻與評分規準

口語測驗因受測者的回答內容為開放性的語言輸出，為避免過於主觀性的評分過程影響了受測者能力判定的結果，需制定一套可靠實用的評分規準。制定評分規準（或稱原則）時，研發人員考量了各等級測驗評量的重點、語言能力表現的特性、語言任務性質的差異等因素，將評分規準的評分重點分為「內容組織」、「表達能力」、「語言運用」三個向度，並在此基礎之上為各題型訂定相應的評分原則。

口語測驗透過標準設定（standard setting）程序，制訂各等級通過門檻。由於本測驗採多元計分制（polytomous items），與單選題非對即錯的概念不同，通過門檻設定方法乃參考 Yes / No Angoff 法（Impara & Plake, 1997）之概念，再因應測驗形式為建構反應題加以調整後施行。所有標準設定成員均由華語文及語言學領域專家所組成，依循標準化流程執行，並在程序性效度與內部效度二項效度證據均獲得支持（國家華語測驗推動工作委員會，2013b；2014；2016）。口語測驗的準備級、入門基礎級各題採 0-3 級分之評分級距，進階高階級與流利精通級各題則均採 0-5 級分之評分級距，各等級通過標準，請見下表 4：

表 4 口語測驗通過分數

測驗等級	通過等級	分數範圍
準備級	準備級一級	16~31分
	準備級二級	32~48分
入門基礎級	入門級	5~9分
	基礎級	10~15分
進階高階級	進階級	12-23分
	高階級	24-30分
流利精通級	流利級	21-34分
	精通級	35分

寫作測驗評分方式對具一定篇幅的段落文本採取分析式評分，111 年至 112 年，寫作測驗將國家教育研究院所研製的《臺灣華語文能力基準》納入參考，更新部分題型，並進行各等級評分規準的調整，研發人員考量各等級寫作任務的差異、篇章組織連貫性、語言文字表現力等因素，將評分原則分為「任務完成度」、「文意連貫性」和「形式語言適切性」三個向度，並根據不同題型與各別試題的特性，在評分原則的基礎上，制定評分細則，並請評分者根據標準評分程序，參照評分原則與細則進行評閱。

同時，因應題型與評分標準的調整，寫作測驗於 111 至 112 年間透過標準設定程序，制訂各等級通過門檻。所有標準設定成員均由華語文及語言學領域專家所組成，依循標準化流程執行，並在程序性效度與內部效度二項效度證據均獲得支持（Lan et al., 2023）。寫作測驗各等級題型則均採 0-5 級分之評分級距，各等級通過標準，請見下表 5。

表 5 寫作測驗通過分數

測驗等級	通過等級	分數範圍
入門基礎級	入門級	4-6分
	基礎級	7-10分
進階高階級	進階級	4-6分
	高階級	7-10分
流利精通級	流利級	5-7分
	精通級	8-10分

二、測驗標準化流程

口語和寫作測驗標準化流程如圖 1 所示，共包含：試題收集、修審、題庫輸入、組合正式卷、評分與成績檢核、考後結果分析等六個步驟，加上對外舉辦的正式考試與考後成績公佈，完成整套「測驗標準化流程」。

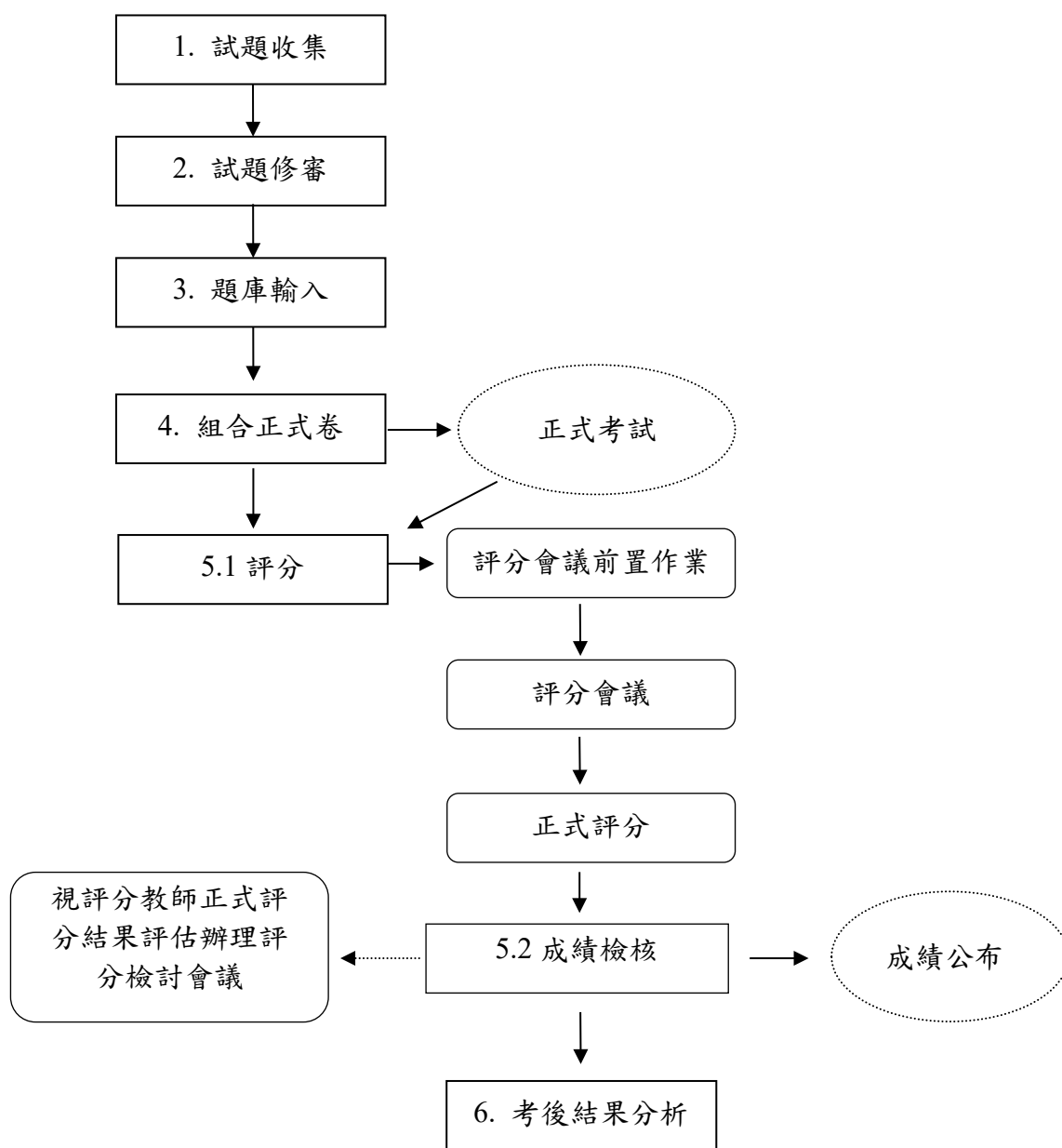


圖 1 測驗標準化流程

口語和寫作測驗「標準化流程」各步驟說明如下表 6。

表 6 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. 組合正式卷	電腦測驗：進行繁、簡版題目檢核；多國語檢核。
5. 評分/成績檢核	5.1 辦理「評分會議」：相關流程詳下表 7 5.2 成績報表檢核：測驗小組成員與統計分析人員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。
6. 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。

口語和寫作測驗「評分流程」各步驟說明如下表 7。

表 7 評分流程說明

項目	說明
評分研習	考前辦理評分研習與培訓：研發人員於每年初辦理評分研習，廣召評分教師進行研習與培訓核心評分教師。
評分會議前置作業	當次考試樣卷挑選會議，建立評分細則。
評分會議	舉辦評分會議，確認評分細則，並進行試評。
正式評分	評分教師正式評分：每份卷子需要 2 位教師評分，如評分差異大於 2 分則需有第 3 人評分。
成績檢核	測驗成績產出，交付統計檢核，確認成績正確，對外公告成績。

參、測驗效能分析

一份測驗是否能夠發揮效用，並能確切地測量受測者的目標潛在能力，通常可通過該測驗的信度與效度分析來進行整體性評估。本節以本年度一至六月入門基礎級、進階高階級與流利精通級之國內正式考試信效度分析結果說明口語測驗及寫作測驗之效能。

一、評分者信度

口語和寫作測驗的信度分析著重於評分者給分是否具有的一致性，亦即評分者信度，此信度可從「評分者間信度」與「評分者內信度」來進行檢驗，而評分者間信度又可從評分者的「嚴格度」與「斯皮爾曼等級相關」進行分析。

口語測驗呈現入門基礎級（3月、5月）、進階高階級（3月）與流利精通級（5月）正式考試的評分結果，兩場入門基礎級考試評分者均為17位，進階高階級和流利精通級則為10位和4位。分析結果顯示，入門基礎級測驗整體評分者嚴格度落在 ± 0.5 logit 以內的百分比分別為88%及94%；進階高階級測驗評分者嚴格度介於 ± 0.5 logit 以內的百分比為80%，表示大多數評分者嚴格度適中。流利精通級測驗則是50%評分者給分較為一致。此外，三個等級評分者之間的斯皮爾曼等級相關係數絕大多數達到高度正相關，整體看來，評分者間信度大致良好，惟流利精通級有待加強。在評分者內一致性方面，絕大多數評分人員皆符合適配度標準，評分者內信度良好。

寫作測驗入門基礎級（3月）、進階高階級（3月）與流利精通級（5月）皆分為二類題型，各題型分別有3到4位評分者參與閱卷工作。寫作測驗評分者信度結果如下：入門基礎級、進階高階級各題型評分者給分的嚴格度介於 ± 0.5 logit 以內的百分比均為100%，顯示評分者給分的嚴格度適中；流利精通級兩個題型評分者給分的嚴格度介於 ± 0.5 logit 之間的百分比為50%和0%，顯示評分者嚴格度不一致，有待加強。

此外，斯皮爾曼等級相關係數方面，進階高階級兩個題型的相關係數大多為達到高度正相關，顯示評分者間一致性大致良好；流利精通級兩個題型的相關係

數為達到中度或高度正相關，顯示評分者間一致性尚可。在評分者內一致性方面，所有評分人員皆符合適配度標準，評分者內信度良好。

為了確保評分教師的評分品質，針對評分結果較差，如偏嚴格、偏寬鬆之評分教師，口語和寫作測驗研發人員皆再進行個別溝通，並提供其評分嚴格度及穩定性的分析結果，做為自我調整之依據，以改善其評分一致性。同時也將這些評分教師列入觀察名單，若後續評分狀況仍未改善，即不續聘。關於5月口語測驗與寫作測驗流利精通級評分者間一致性略低的狀況，經檢討可能肇因於考生人數較少，且部分評分人員的訓練可再加強，研發人員已執行後續改善方案，詳細處理方式，將於下文中詳述。

（一）評分者間信度（嚴格度分析）

以軟體 Facets¹部分給分模式（Partial Credit Model，以下簡稱 PCM）²對資料進行多面向 Rasch 分析（many facet Rasch measurement，以下簡稱 MFRM）³。

口語測驗方面，入門基礎級、進階高階級、流利精通級口語測驗評分的嚴格度結果，分別如表 8 至表 11 所示。以嚴格度介於 ± 0.5 logit 以內為標準，入門基礎級口語測驗3月場次的17位評分者中，除了 A50、A05 給分較為寬鬆（-0.54 與 -1.00）外，其餘 15 位評分者嚴格度均落在 ± 0.5 logit 以內，百分比為 88%；5月場次有 17 位評分者，其中 16 位嚴格度落在 ± 0.5 logit 以內，94%的評分者給分相近，僅 A05 給分較為寬鬆（-0.68）。進階高階級測驗方面，10 位評分者中有 8 位評分者嚴格度介於 ± 0.5 logit 以內，百分比為 80%，B39 較為嚴格（0.96），而 B04 略為寬鬆（-0.53）。流利精通級測驗四位評分者中有兩位嚴格度落在 ± 0.5 logit 以內，給分較寬鬆和較嚴格的各有一位，50%評分者嚴格度相近。整體而言，入門基礎級和進階高階級測驗的評分者間信度大致良好，而流利精通級評分者間信度有待改善。

對此，已於 6 月底至 7 月中針對給分較寬鬆及較嚴格的評分者規劃兩個輔導

¹ Linacre, J. M. (2013). Facets® (Version 3.71.3)[Computer Software]. Beaverton, Oregon: Winsteps.com.

² Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. MESA Press.

³ Linacre, J. M. (1988). *Many-facet Rasch measurement*. MESA Press.

機制，首先個別溝通評分偏誤造成的原因，並透過線上回訓活動讓評分者調整評分技巧；後續則會在此類評分者參與下次評分工作前的定期回訓過程中，增加回訓音檔的練習數量。

表 8 入門基礎級口語測驗評分者嚴格度（3 月）

評分者	評閱音檔數	觀察的平均值	調整過平均值	嚴格度	標準誤	INFIT MNSQ
A24	165	1.6	1.20	0.46	0.16	1.04
A22	170	1.5	1.24	0.34	0.16	1.29
A26	165	1.6	1.24	0.34	0.16	1.35
A20	165	1.6	1.25	0.29	0.16	1.10
A36	170	1.4	1.26	0.26	0.15	1.03
A63	170	1.1	1.29	0.17	0.16	1.54
A02	170	1.3	1.31	0.10	0.15	0.97
A41	170	1.5	1.31	0.10	0.16	0.91
A31	170	1.1	1.32	0.07	0.16	1.09
A46	170	1.3	1.33	0.04	0.15	1.17
A62	170	1.5	1.35	0.01	0.15	1.10
A64	170	1.2	1.35	-0.01	0.16	0.91
A16	131	1.5	1.38	-0.11	0.18	0.90
A65	170	1.2	1.40	-0.16	0.16	1.22
A18	170	1.6	1.46	-0.36	0.16	1.12
A50	170	1.5	1.52	-0.54	0.15	1.09
A05	170	1.8	1.66	-1.00	0.15	1.39

表 9 入門基礎級口語測驗評分者嚴格度（5月）

評分者	評閱 音檔數	觀察的 平均值	調整過 平均值	嚴格度	標準誤	INFIT MNSQ
A62	120	1.7	1.72	0.42	0.17	1.02
A36	120	1.6	1.78	0.19	0.17	1.00
A16	240	1.7	1.78	0.16	0.12	0.87
A31	120	1.7	1.79	0.15	0.17	0.79
A22	117	1.7	1.80	0.14	0.17	1.08
A23	115	1.7	1.80	0.11	0.18	0.98
A26	120	1.6	1.81	0.08	0.17	1.12
A41	120	1.8	1.81	0.08	0.18	0.76
A54	120	1.8	1.81	0.06	0.17	1.05
A64	120	1.8	1.84	-0.02	0.18	0.90
A46	120	1.8	1.84	-0.02	0.17	1.02
A20	120	1.8	1.85	-0.06	0.17	1.19
A02	120	1.8	1.86	-0.12	0.17	0.92
A43	120	1.7	1.86	-0.12	0.17	1.24
A53	120	1.8	1.87	-0.17	0.17	1.04
A65	115	1.8	1.88	-0.20	0.18	1.00
A05	115	1.9	2.00	-0.68	0.18	1.42

表 10 進階高階級口語測驗評分者嚴格度

評分者	評閱 音檔數	觀察的 平均值	調整過 平均值	嚴格度	標準誤	INFIT MNSQ
B39	187	1.90	1.62	0.96	0.13	0.96
B51	183	2.30	1.93	0.26	0.12	1.02
B47	194	2.20	1.98	0.17	0.12	0.94
B43	194	2.40	2.03	0.06	0.12	0.86
B48	194	2.40	2.04	0.04	0.12	0.84
B07	189	2.40	2.10	-0.08	0.12	1.25
B24	194	2.40	2.13	-0.16	0.12	0.89
B16	450	2.30	2.21	-0.30	0.08	0.74
B50	187	2.30	2.27	-0.42	0.13	1.38
B04	187	2.40	2.32	-0.53	0.13	1.19

表 11 流利精通級口語測驗評分者嚴格度

評分者	評閱 音檔數	觀察的 平均值	調整過 平均值	嚴格度	標準誤	INFIT MNSQ
C02	231	1.90	1.90	0.68	0.10	1.24
C08	231	2.10	2.15	0.15	0.10	1.25
C06	371	2.30	2.33	-0.25	0.08	0.67
C07	294	2.60	2.50	-0.59	0.09	0.96

寫作測驗方面，評分者在入門基礎級、進階高階級、流利精通級各題型向度分數給分的嚴格度結果呈現於表 12 至表 17 以嚴格度介於 ± 0.5 logit 以內為標準，入門基礎級圖片描述與書信寫作各三名評分者的嚴格度均介於 ± 0.5 logit 之間，顯示所有評分者嚴格度相近，百分比達 100%。進階高階級測驗，書信寫作與觀點論述各四名評分者的嚴格度分布落在 ± 0.5 logit 範圍以內的百分比同樣為 100%，顯示評分嚴格度一致。流利精通級測驗方面，摘要寫作題型有四名評分者進行閱卷，嚴格度落在 ± 0.5 logit 範圍以內的百分比為 50%，C26 較嚴格 (0.54)、C28 較寬鬆 (-1.09)；觀點論述題型有三名評分者進行閱卷，C26 和 C30 嚴格度比較接近，而 C30 較為寬鬆，以致於沒有評分者嚴格度落在 ± 0.5 logit 以內。整體而言，寫作測驗入門基礎級與進階高階級評分者間信度良好，而流利精通級兩個題型的評分者一致性不盡理想。

113 年 5 月流利精通級寫作測驗為該等級首場採用調整後評分規準的正式考試，分析顯示兩題型評分者的評分品質有待加強，可能原因為此次考生文本數量較少，且 4 月針對新規準辦理評分人員招募會議時，首次合作的評分人員較多，因此，後續將持續觀察流利精通級各級分考生表現的典型樣態，使評分招募會議中的訓練內容更為完善，並再次針對該等級辦理評分培訓招募、重新訓練合格之評分人員，以期提升評分品質。

表 12 入門基礎級寫作測驗圖片描述題型評分者嚴格度

評分者	評閱 篇數	觀察的 平均值	調整後 平均值	嚴格度	標準誤	INFIT MNSQ
A43	16	4.0	4.04	0.15	0.34	0.87
A41	16	4.1	4.08	-0.08	0.34	1.13
A36	16	4.1	4.08	-0.08	0.34	0.84

表 13 入門基礎級寫作測驗書信寫作題型評分者嚴格度

評分者	評閱 篇數	觀察的 平均值	調整後 平均值	嚴格度	標準誤	INFIT MNSQ
A44	16	4.00	4.04	0.41	0.29	1.18
A36	16	4.20	4.27	-0.02	0.30	0.73
A37	16	4.30	4.43	-0.38	0.31	0.79

表 14 進階高階級寫作測驗書信寫作題型評分者嚴格度

評分者	評閱 篇數	觀察的 平均值	調整後 平均值	嚴格度	標準誤	INFIT MNSQ
B52	33	3.4	3.31	0.16	0.17	0.99
B07	33	3.5	3.36	0.04	0.17	1.17
B25	33	3.5	3.40	-0.08	0.17	1.15
B16	33	3.5	3.41	-0.11	0.17	0.67

表 15 進階高階級寫作測驗觀點論述題型評分者嚴格度

評分者	評閱 篇數	觀察的 平均值	調整後 平均值	嚴格度	標準誤	INFIT MNSQ
B43	31	2.6	2.5	0.49	0.17	1.31
B16	31	2.8	2.8	-0.02	0.16	0.65
B48	31	2.8	2.8	-0.09	0.16	1.31
B14	31	2.9	3.0	-0.37	0.16	0.67

表 16 流利精通級寫作測驗摘要寫作題型評分者嚴格度

評分者	評閱 篇數	觀察的 平均值	調整後 平均值	嚴格度	標準誤	INFIT MNSQ
C26	41	1.2	1.01	0.54	0.26	0.59
C29	39	1.3	1.01	0.47	0.26	1.46
C27	39	1.3	1.02	0.08	0.25	1.00
C28	39	1.5	1.05	-1.09	0.22	1.13

表 17 流利精通級寫作測驗觀點論述題型評分者嚴格度

評分者	評閱 篇數	觀察的 平均值	調整後 平均值	嚴格度	標準誤	INFIT MNSQ
C26	41	1.4	1.08	0.60	0.20	0.70
C31	39	1.4	1.09	0.52	0.20	0.95
C30	39	1.9	1.34	-1.13	0.16	1.24

(二) 評分者間信度 (斯皮爾曼等級相關)

口語測驗針對入門基礎級、進階高階級與流利精通級測驗評分結果進行斯皮爾曼等級相關分析，以了解不同等級各組評分者兩兩之間的評分者間信度。結果如表 18 所示，因篇幅有限，僅呈現各組別內兩兩評分者之間相關係數的整體平均值。一般來說，相關係數達.40 以上表示有中度相關，達.70 以上則表示有高度相關。入門基礎級 3 月正式考試分為 5 組，所有組別評分者相關係數平均值達.80 以上，具有高度正相關；5 月正式考試分為 5 組，所有組別相關係數平均值達.70

以上；進階高階級正式考試4組評分者的斯皮爾曼等級相關係數的平均值都達.80以上，具高度正相關；流利精通級正式考試分為2組，兩組給分的相關係數平均值分別為.54和.87。

整體來看，入門基礎級與進階高階級測驗，兩兩評分者之間給分均有高度正相關存在，評分者間信度良好。流利精通級測驗方面，則是有一組兩位評分者間信度不盡理想⁴。但華測會訂有成績處理流程，若同一位考生的評分結果不一致時，將交由核心評分者評定成績，以確保考生成績不會受到評分者嚴格度不一致所影響。此外，為避免同一組評分人員評分嚴格度同時偏寬鬆或偏嚴格，研發人員會依據評分人員過去評分表現進行分組，確保每一組別至少有一名嚴格度良好、評分一致性較佳之評分人員。

表 18 口語測驗評分者間斯皮爾曼等級相關

測驗等級	組別					整體 平均值
	1	2	3	4	5	
入門基礎級（3月）	.82	.89	.84	.89	.83	.85
入門基礎級（5月）	.80	.78	.83	.84	.84	.82
進階高階級	.85	.90	.89	.83		.87
流利精通級	.54	.87				.70

寫作測驗針對進階高階級與流利精通級測驗評分結果進行斯皮爾曼等級相關分析，因各等級測驗評分組別均為一組，在此呈現組內評分者兩兩之間相關係數的平均值，結果如表 19 所示。進階高階級書信寫作題型的整體級分方面，評分者等級相關係數平均值為.78，具高度正相關。評分向度方面，「任務完成度」和「形式語言適切性」相關係數平均值也高於.70，具高度正相關，「文意連貫性」略低，平均值為.64。觀點論述題型的整體級分平均值為.81，達高度正相關；評分向度方面，同樣為「任務完成度」和「形式語言適切性」相關係數平均值高於.70，有高度正相關，而「文意連貫性」平均值低於.70。

在流利精通級摘要寫作題型的整體級分方面，評分者相關係數平均值為.74，達高度正相關。評分向度方面，「任務完成度」和「形式語言適切性」相關係數

⁴ 流利精通級相關係數較低的狀況，已於當年度6月至7月之評分人員回訓時進行檢討。

平均值也高於.70，具高度正相關，「文意連貫性」略低，平均值為.64。至於觀點論述題型的整體級分、三個評分向度級分的等級相關係數平均值則均低於.70，達到中度正相關。

整體來看，進階高階級測驗兩種題型和流利精通級摘要寫作題型，評分者之間給分大多有正相關存在，僅一個分向度未達到.70 以上的高度相關，評分者間信度大致良好，「文意連貫性」給分的一致性待加強。至於流利精通級觀點論述題型，整體級分和向度級分的相關分析結果均不盡理想，後續如上節所述，將重新培訓流利精通級評分人員，以期加強評分人員對新規準的認知與掌握，提升評分一致性。然而，華測會訂有成績處理流程，若同一位考生的評分結果不一致時，將交由核心評分者評定成績，確保考生成績不會受到評分者嚴格度不一致所影響。

表 19 寫作測驗評分者間斯皮爾曼等級相關

測驗等級	題型	向度（細項）	平均值
進階 高階級	書信寫作	整體級分	.78
		任務完成度	.75
		文意連貫性	.64
		形式語言適切性	.77
	觀點論述	整體級分	.81
		任務完成度	.85
		文意連貫性	.66
		形式語言適切性	.74
流利 精通級	摘要寫作	整體級分	.74
		任務完成度	.75
		文意連貫性	.64
		形式語言適切性	.75
	觀點論述	整體級分	.61
		任務完成度	.54
		文意連貫性	.67
		形式語言適切性	.69

註：因入門基礎級測驗扣除樣卷、標準卷後有效樣本數未達 20 人，未進行分析。

（三）評分者內信度

此節透過嚴格度分析中評分者之嚴格度標準誤以及訊息加權均方差適配指

標 (information-weighted mean-square fit statistic, 以下簡稱 INFIT MNSQ) 數值檢視評分者本身給分的一致性情形。

口語測驗方面，入門基礎級分析結果如表 8 與表 9 所示，3 月場次評分者 A16 由於為核心評分者，評閱音檔數較少，故標準誤較大，排除 A16 後，各評分者之標準誤介於 0.15 至 0.16 之間；5 月場次 A16 同時協助挑選樣本音檔，評閱音檔數較多，標準誤較小，排除後其餘評分者標準誤介於 0.17 至 0.18 之間，兩次考試的標準誤差異均不大，顯示各評分者的評分穩定度佳。另一檢視標準——適配性指標需介於 0.50 至 1.50 之間，兩次考試中僅 3 月有 1 位評分者 (A63) 略高於 1.50，其餘評分教師的表現均符合指標，顯示評分者內一致性佳，自身評分穩定性良好。進階高階級結果如表 10 所示，排除 B16 為核心評分人員協助挑選樣本音檔故評閱數量較高，標準誤較小外，各評分者之標準誤介於 0.12 至 0.13，評分者的評分穩定度佳；適配性指標部分，所有評分教師的表現均符合指標，評分穩定性良好。至於流利精通級結果，從表 11 可知，扣除核心評分者 C06，其餘三位評分者標準誤相近，介於 .09 至 .10 之間，且適配性指標數值均符合標準，表示評分者自身穩定性良好。

寫作測驗方面，入門基礎級圖片描述與書信寫作分析結果如表 12、表 13 所示，圖片描述三名評分者標準誤均為 0.34，書信寫作 3 名評分者則為 0.29、0.30、0.31，差異均不大，表示各評分者給分穩定度佳；所有評分者的適配性指標均介於 0.50 至 1.50 之間，亦表示評分者內一致性佳，自身評分穩定性良好。進階高階級兩個題型的結果見表 14 及表 15，各個評分者標準誤均為 0.17，以及 0.16 至 0.17 之間，差異皆不大；所有評分者的適配性指標數值均符合標準。至於流利精通級兩個題型分析結果如表 16、表 17 所示，書信寫作四名評分者標準誤介於 0.22 至 0.26 之間，觀點論述三名評分者標準誤為 0.20、0.20 和 0.16；所有評分者的適配性指標數值同樣也符合標準，評分者內一致性佳。

為了確保評分教師的評分品質，口語和寫作測驗皆個別提供各評分教師其自身評分嚴格度及穩定性的統計分析結果，做為自我調整改善評分品質的參考依據，使評分教師能更加掌握評分規準，給予受測者更為客觀、合理、適切的成績，

避免未來再度出現評分過度嚴格或寬鬆的情況，改善其評分一致性。同時也會將評分嚴格度差異較大的評分教師列入觀察名單，若後續評分狀況仍未改善，即不續聘。

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，口語、寫作測驗效度分析藉由試題分析、因素分析，評估測驗的建構效度，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。此外，口語和寫作測驗也藉由固定、標準化的評分流程說明程序性效度。

（一）建構效度

由 IRT 試題分析與驗證性因素分析結果可知，華語文口語、寫作測驗各等級皆具有一定之建構效度。

1. IRT 試題分析

口語和寫作測驗因採用多元計分，以分析軟體 Facets 的 PCM 模式對資料進行 MFRM 分析，檢視試題難度參數資料和模式適配情形。考量考生成績易受到評分者因素影響，一般採用 INFIT MNSQ 介於 0.50 到 1.50 以及 INFIT ZSTD 介於 -3.00 到 3.00 的標準評估試題是否與模式適配，表 20 的分析結果顯示口語測驗入門基礎級 5 月正式考試、進階高階級、流利精通級各一場正式考試試題適配率均為 100%，僅入門基礎級 3 月正式考試有 1 題些微不適配（INFIT MNSQ 為 1.51、INFIT ZSTD 為 6.4），顯示幾乎所有試題測量到相同構念的華語文口語能力。

寫作測驗方面，進階高階級書信寫作和觀點論述題型，均分為任務完成度、文意連貫性與形式語言適切性 3 個評分向度，兩個題型的適配率均為 100%；流利精通級摘要寫作和觀點論述題型為同樣分為 3 個評分向度，兩個題型的適配率也都為 100%，表示所有向度測量到相同構念的華語文寫作能力。

表 20 口語測驗、寫作測驗試題適配分布

測驗類型	測驗等級	總題數/細項	適配題數	適配率
口語	入門基礎級(3月)	5	4	80%
	入門基礎級(5月)	5	5	100%
	進階高階級	6	6	100%
	流利精通級	7	7	100%
寫作	進階高階級書信寫作	3	3	100%
	進階高階級觀點論述	3	3	100%
	流利精通級摘要寫作	3	3	100%
	流利精通級觀點論述	3	3	100%

註：入門基礎級寫作測驗因扣除樣卷、標準卷後有效樣本數未達 20 人，未進行分析。

2. 驗證性因素分析

此節使用 Mplus⁵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（國家華語測驗推動工作委員會，2013a）。口語測驗各等級的因素結構均採單因素模式，亦即口語表達能力；寫作測驗各亦採用單因素模式來測量寫作表達能力⁶。由於篇幅有限，在此僅呈現 3 月入門基礎級口語測驗和進階高階級寫作測驗單因素模式結果，其餘測驗等級的正式考試因素負荷量等參數請見附件 1。

3 月入門基礎級口語測驗正式考試驗證性因素分析結果如圖 2 所示，在基本適配指標部分，單因素模式驗證性因素分析結果顯示，試題因素負荷量介於.85 至.97 之間，各題因素負荷量統計考驗均達顯著水準 ($p<.05$)。進階高階級寫作測驗單因素模式分析結果如圖 3 所示，兩個題型各評分向度的因素負荷量介於.70 至.96 之間，因素負荷量統計考驗均達顯著水準 ($p<.05$)。

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

⁶ 寫作測驗雖包含兩個題型，各具三個評分向度，理論上可建構雙因素模型。然而，由於考生人數較少，低於一般驗證性因素分析建議的穩健樣本量（大於 200 人），為確保模型估計的穩定性與收斂性，並避免小樣本導致的參數不確定性，本研究地採用參數更精簡、自由度較大的單因素模型進行結構效度檢驗。

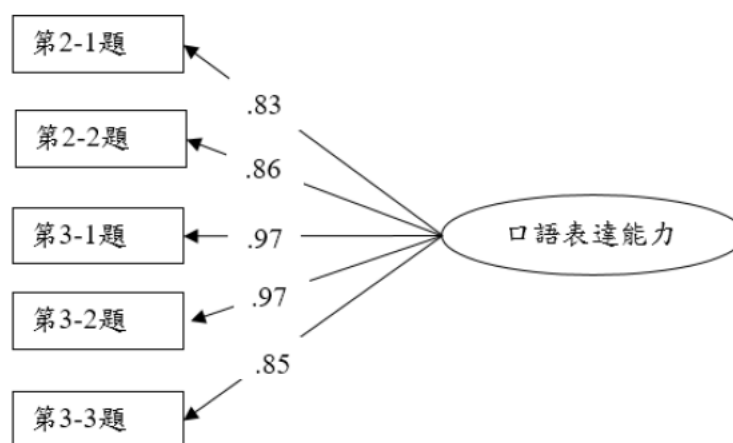


圖 2 入門基礎級口語測驗單因素模式 (3 月)

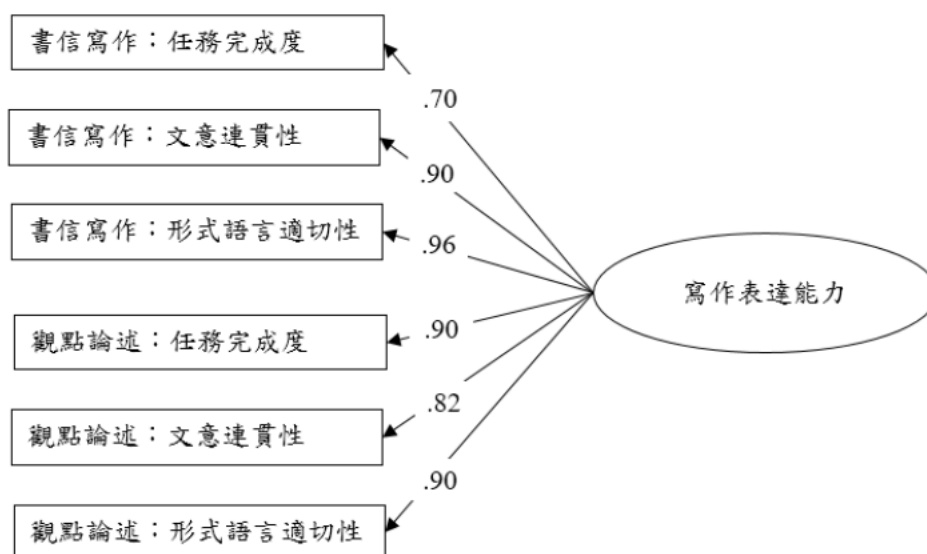


圖 3 進階高階級寫作測驗單因素模式

透過整體模式適配度分析，可評估理論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) 評估模式適配度，一般建議應小於 3，由表 21 可知，口語和寫作測驗的單因素模型多數卡方自由度比均大於 3；而絕對適配度評估標準為平方概似平方誤根係數 (root mean square error of approximation，以下簡稱 RMSEA) 小於 .08，下表多數模式的 RMSEA 都大於 .08。不過，文獻指出，由於

χ^2 統計量在小樣本與高自由度模型下容易膨脹，因此 χ^2/df 指標可能高估模型不適配程度 (Kenny & McCoach, 2003)；RMSEA 指標也在小樣本時有高估的現象，使適配模式被視為不理想模型 (Bentler & Yuan, 1999，引自邱皓政，2011)。考量到口語和寫作測驗各個模式的考生人數均小於或接近 100 人，前述兩指標數值偏高可能是受人數影響，需搭配其他指標進行評估。

在增值適配度評估方面，口語、寫作測驗各等級的比較適配指標 (comparative-fit index，簡稱 CFI) 和非規範適配指標 (non-normed fit index，簡稱 NNFI) 數值均大於 .90，顯示模式均符合增值適配度指標。此外，各模式中觀察變項的標準化因素負荷量均接近或大於 0.70 (見附件 1)，顯示變項與潛在能力之間關連强度高。因此，研究者將主要判斷依據聚焦在較為穩定的 CFI 與 TLI 指標，並結合良好的因素負荷量，評估口語和寫作測驗的單因素模型建構效度獲得支持。

表 21 口語測驗、寫作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
口語	入門基礎級單因素模式 (3 月)	2.37	0.12	0.998	0.996
	入門基礎級單因素模式 (5 月)	4.11	0.24	0.972	0.944
	進階高階級單因素模式	2.31	0.13	0.998	0.997
	流利精通級單因素模式	1.32	0.08	0.999	0.998
寫作	進階高階級單因素模式	5.90	0.37	0.956	0.926
	流利精通級單因素模式	3.54	0.25	0.989	0.981

註：寫作測驗入門基礎級因考生人數未達 30 人，未進行分析。

(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，口語和寫作測驗各等級正式考試結束後，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，口語測驗各等級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；寫作測驗進階高階級和流利精通級整體自評與測驗成績相關分析結果未達顯著水準，有待持續收集效度證

據。

口語測驗方面，入門基礎級兩次正式考試考生自評結果總和與測驗總分的積差相關係數分別為.45 ($p < .01$) 與.30 ($p < .05$)；而進階高階級、流利精通級考生自評與測驗總分的積差相關係數分別為.65 以及.39，均達到.01 顯著水準，顯示各個測驗等級考生自評口語能力與測驗總分之間有正相關存在，自評口語能力越佳者，其口語測驗總分越高，具有良好效度。

將自評問卷各題回答結果與測驗總分進行斯皮爾曼等級相關分析，結果如表 22 所示，入門基礎級 3 月場次 13 題自評問題中，只有第 12 題相關係數未達.05 顯著水準，其餘 12 題與測驗總分相關係數介於.26 至.43 之間，達到.05 或.01 顯著水準，表示在這些題目中自評能力越高的考生，其測驗總分也越高，顯示出良好的效標關聯效度。至於 5 月場次自評問卷與測驗表現之間的相關係數普遍較低，然而過去考生人數達 30 人以上之正式考試分析結果，相關係數大多至少可達.05 顯著水準。推測可能原因為此場次考生成績表現普遍相當理想，整體獲證百分比高達 88.5%，填答問卷的考生中，未通過者僅有個位數，因群體同質性較高使得相關係數降低。

進階高階級 13 題的自評問題與測驗總分相關係數介於.44 至.61 之間，均達到.01 顯著水準，表示在這些題目中自評能力越高的考生，其測驗總分也越高，顯示出良好的效標關聯效度。流利精通級部分，八題自評問題中，第 5 題、第 7 題、第 8 題與測驗總分相關係數未達.05 顯著水準，其餘問題相關係數介於.28 至.44 之間，均達到顯著水準 ($p < .05$)，效標關聯效度尚可。詳細問卷內容請見附件 2。針對相關分析結果不理想之問卷題目，將持續觀察，若連續兩次分析結果均未達.05 顯著水準，則考慮修改內容或刪除。

表 22 口語測驗自評問卷各題與測驗總分之相關分析

入門基礎級 (3 月) (N=76)	Q1	Q2	Q3	Q4	Q5	Q6	Q7
	.26*	.28*	.41**	.34**	.28*	.43**	.30**
	Q8	Q9	Q10	Q11	Q12	Q13	
	.36**	.34**	.40**	.28*	0.20	.28*	
入門基礎級 (5 月) (N=51)	Q1	Q2	Q3	Q4	Q5	Q6	Q7
	.11	.05	.36**	.31*	.24	.24	.20
	Q8	Q9	Q10	Q11	Q12	Q13	
	.17	.20	.13	.20	.24	.26	
進階高階級 (N=46)	Q1	Q2	Q3	Q4	Q5	Q6	Q7
	.61**	.54**	.47**	.53**	.56**	.51**	.48**
	Q8	Q9	Q10	Q11	Q12	Q13	
	.61**	.44**	.52**	.50**	.55**	.45**	
流利精通級 (N=53)	Q1	Q2	Q3	Q4	Q5	Q6	Q7
	.33*	.44**	.28*	.32*	.13	.34*	.21
	Q8						
	.18						

註： * $p < .05$ ； ** $p < .01$ 。

寫作測驗方面，流利精通級考生自評結果與測驗總分的積差相關係數為.26 ($p > .05$)，顯示考生自評寫作能力與測驗總分之間沒有正相關存在。如表 23，六題自評問題中僅第一題與測驗總分相關係數為.35，達到.05 顯著水準，其餘五題均未達到顯著水準 ($p > .05$)，檢視各題回答選項與考生通過等級交叉分析表，發現本次流利級精通級考生測驗表現普遍不佳，通過百分比比較低 (7.3%)，且未通過之考生，在各題均以回答「能做到」或「大概能做到」占多數 (回答「能做到」百分比介於 44.7%至 55.3%之間，「能做到」或「大概能做到」兩選項合計 84.2%至 92.1 之間)，僅極少數回答「不太能做到」，以致於相關係數偏低。由於問卷內容在去年已依據過去分析結果進行修改，將持續觀察以評估是否再做調整。詳細問卷內容請見附件 3。

表 23 寫作測驗自評問卷各題與測驗總分之相關分析

流利精通級 (N=41)	Q1	Q2	Q3	Q4	Q5	Q6
	.35*	.15	.13	.15	.27	.26

註 1：* $p < .05$ 。

註 2：入門基礎級、進階高階級測驗因有效填答人數未達 30 人，樣本代表性較不足，故未進行分析。

(三) 程序性效度

口語和寫作測驗是一種表現測驗 (Performance assessment)，考生的成績由評分者根據評分原則來評定，而評分涉及主觀判斷，若未確實掌握評分標準，將無法正確區分考生能力，亦連帶影響測驗效度，因此評分者訓練至關重要。在此依序陳述口語和寫作測驗辦理評分活動的場次、評分訓練的流程及訓練後正式評閱的結果。由評分者問卷分析結果顯示，評分者均同意會議帶領者對於評分原則、標準音檔/文本等資料的內容說明得很清楚，評分者對於自己評定的成績有信心，加上每場次評分會議均依照固定程序辦理，口語和寫作測驗具有程序性效度。

口語測驗與寫作測驗 113 年度為維持評分品質，配合入門基礎級、進階高階級與流利精通級正式考試的實施，皆辦理相應之評分培訓與評分會議，確保參與正式考試之評分者，皆在正式評分前接受完整訓練

為增加熟悉新評分規準之流利精通級儲備評分教師人數，寫作測驗研發人員針對新評分規準，113 年於 5 月舉辦 1 場流利精通級儲備教師評分研習，共 14 名教師參加。在評分的作業流程上，為確保評分品質，口語和寫作測驗均制定一套固定的流程並在每次正式考試中執行。程序包括：正式評分會議前，讓評分人員進行試評，調校其評分標準，之後評分人員獨立進行正式評分工作。正式評分工作結束之後，研發人員彙整成績送交統計分析人員，進行評分嚴格度、評分者間與評分者內一致性分析，以了解評分人員評分品質，並將結果回饋給評分人員。詳細評分流程請見表 24 與 25。

表 24 口語測驗評分會議流程

階段	項目	內容
評閱前	評分人員培訓	1. 挑選樣本音檔、製作評分回饋樣本。 2. 線上實作練習、個別提供評分回饋。 3. 重新校正評分。
	評分會議準備	1. 制定會議流程。 2. 篩選並確認待擬細則之音檔。 3. 挑選共同題各級分樣本音檔，設定評分系統。
評閱中	評分會議	1. 說明評分流程。 2. 與評分人員共擬評分細則。 3. 現場試評，討論評分結果，建立評分共識。
	正式評分	1. 進行為期二週的評分工作。 2. 監控評分人員評分情況。
評閱後	評分結果彙整及分析	1. 彙整、提交評分結果，分析評分嚴格度、評分者間與評分者內一致性。 2. 據分析結果與評分人員進行溝通及再培訓。

表 25 寫作測驗評分作業流程

階段	項目	內容
評閱前	評分人員培訓	1. 挑選評閱卷。 2. 線上實作練習、個別提供評分回饋。 3. 重新校正評分。
	評分會議準備	1. 制定會議流程。 2. 依據試題制定評分細則。 3. 挑選各級樣卷、標準卷、練習卷，匯入評分系統。
評閱中	評分會議	1. 說明評閱流程。 2. 說明評閱重點、評分規準與細則。 3. 監控評分人員評分一致性情況。
	正式評分	1. 進行為期二週的評分工作。 2. 監控評分人員評分情況。
評閱後	評分結果彙整及分析	1. 彙整、提交評分結果，分析評分嚴格度、評分者間與評分者內一致性。 2. 據分析結果與評分人員進行溝通及再培訓。

除每場次評分工作皆遵循前述之標準化流程辦理外，華測會另於評分會議結

束後對評分人員進行問卷調查，了解評分人員對於評分內容的理解程度，以及自評給分的信心程度等等。在此因篇幅有限，呈現入門基礎級口語測驗與寫作測驗問卷調查結果，其餘測驗等級問卷結果請參見附件 4。

從表 26 與表 27 可知，入門基礎級口語測驗和寫作測驗的評分人員問卷調查結果，除了口語測驗第 8 題的兩場測驗的同意百分比均為 86% 外，其餘各題同意百分比達 96% 以上。此結果顯現評分人員理解會議帶領者對於評分原則、標準音檔/文本等資料的說明，會議中提供的相關文件、細則討論與試評過程也有助於其進行正式給分；評分人員大致上對自己的給分具有信心。以上亦顯示評分會議的辦理流程具有程序性效度。

表 26 入門基礎級口語測驗評分人員問卷調查結果

問卷題目	3月		5月	
	平均數	同意百分比	平均數	同意百分比
1. 評分會議中，會議帶領者對於音檔試評及試評結果討論的形式，說明得很清楚。	5.0	100.0%	4.9	98.0%
2. 評分會議中，會議帶領者對於評分原則、標準音檔及能力指標的內容解說清楚。	5.0	100.0%	4.9	98.0%
3. 承上，評分原則、標準音檔及能力指標的解說，有助於我了解各等級使用者的口語表現。	5.0	100.0%	4.9	98.0%
4. 評分會議中，會議帶領者對於評分原則、標準音檔及能力指標如何應用在評分過程，說明得很清楚。	5.0	100.0%	4.9	98.0%
5. 承上，評分原則、標準音檔及能力指標應用技巧的解說，有助於我進行評分。	5.0	100.0%	4.9	98.0%
6. 評分會議中，試評音檔並就試評結果進行討論，有助於我掌握評分原則以進行評分。	5.0	100.0%	4.9	98.0%
7. 評分會議中，共同討論細則，有助於我掌握評分原則進行評分。	5.0	100.0%	4.8	96.0%
8. 我對於自己在評分階段所評定的音檔成績及評分依據有信心。	4.3	86.0%	4.3	86.0%

註：問卷填答方式，「1」表示非常不同意，「2」表示不同意，「3」表示普通，「4」表示同意，「5」表示非常同意。

表 27 入門基礎級寫作測驗評分人員問卷調查結果

問卷題目	平均數	同意百分比
1. 我了解此次華測會舉辦【評分會議】的目的。	4.0	100.0%
2. 會議帶領者對於評分原則與細則說明得很清楚、詳盡。	4.0	100.0%
3. 會議帶領者進行練習卷評閱及團體討論，讓我更清楚掌握評分原則與細則，並且有助於正式評閱工作。	4.0	100.0%
4. 我根據評分原則與細則進行正式評閱工作。	4.0	100.0%
5. 我對於自己在此次正式評閱中所評定的成績有信心。	4.0	100.0%

註：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。

肆、結論

本文主要針對華語文口語和寫作測驗之能力描述、測驗題型與題數、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的信度與效度，以評估測驗效能。

在測驗信度分析方面，由於本測驗之受測者成績主要仰賴評分者判定，因此，受測者成績除了受到受測者自身具備之口語或寫作能力與測驗試題難度的影響之外，亦會受到評分者嚴格度變異的影響，故本測驗主要以「評分者間給分一致性」與「評分者自身給分穩定性」二個面向評估測驗信度。結果顯示，口語測驗入門基礎級和進階高階級評分者間信度大致良好，而流利精通級尚待加強；評分者內信度則是所有等級都良好。寫作測驗部分，入門基礎級和進階高階級的評分者間信度大致良好，流利精通級同樣需加強培訓；而三個等級的評分者內信度皆良好。針對評分嚴格度變異較大的評分者，本會將進行評分再訓練並列入觀察名單，若日後評分結果未獲改善，即不予續聘。

在測驗效度分析部分，進行了試題分析與驗證性因素分析，主要目的在於檢視受測者測驗表現是否與口語和寫作測驗制訂的測量架構相符，以此作為建構效度的證據。結果顯示，兩項測驗各等級的單因素模式皆具有建構效度。此外，透過受測者自評結果與測驗成績的相關分析，評估測驗的效標關聯效度。分析結果顯示，口語測驗具備效標效度證據，而寫作測驗本年度的效標關聯效度結果不盡理想，本會將針對問卷內容進行調整，持續收集效度證據。而為使評分者有效遵循測驗所擬定之評分原則，並給予適切評分，本會採取標準化的評分流程來培訓評分者，並以問卷調查評分者對於評分會議的掌握程度，提供了程序性效度的證據。

伍、參考文獻

- 邱皓政 (2011)。結構方程模式：LISREL / SIMPLIS 原理與應用。雙葉書廊。
- 國家華語測驗推動工作委員會 (2013a)。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。
- 國家華語測驗推動工作委員會 (2013b)。華語文能力測驗技術報告—2013 (3) 口語測驗信效度。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 (3) 口語測驗信效度。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 (2016)。華語文能力測驗技術報告—2016 (1) 口語測驗信效度。國家華語測驗推動工作委員會。
- Impara, J. C., & Plake, B. S. (1997). Standard-setting: An alternative approach. *Journal of Educational Measurement*, 34, 353–366.
- Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 10(3), 333–351.
- Lan, P.-J., Chen, P.-H., Peng, S.-H., Chao, C.-P., Wang, T.-A., & Tseng, W.-H. (2023). *Aligning ACTFL writing proficiency guidelines with CEFR descriptors: Insights from Chinese writing assessment* [Manuscript submitted for publication]. Steering Committee for the Test of Proficiency-Huayu.

陸、附件

附件 1 口語、寫作測驗正式考試驗證性因素分析結果

表 1-1 入門基礎級口語測驗（3 月）各題因素負荷量摘要表

試題編號	因素負荷量	標準誤	P 值
第 2-1 題	0.83	0.04	0.00
第 2-2 題	0.86	0.03	0.00
第 3-1 題	0.97	0.02	0.00
第 3-2 題	0.97	0.02	0.00
第 3-3 題	0.85	0.03	0.00

表 1-2 入門基礎級口語測驗（5 月）各題因素負荷量摘要表

試題編號	因素負荷量	標準誤	P 值
第 2-1 題	0.79	0.07	0.00
第 2-2 題	0.69	0.09	0.00
第 3-1 題	0.98	0.06	0.00
第 3-2 題	0.82	0.04	0.00
第 3-3 題	0.79	0.06	0.00

表 1-3 進階高階級口語測驗各題因素負荷量摘要表

試題編號	因素負荷量	標準誤	P 值
第 2-1 題	0.92	0.03	0.00
第 2-2 題	0.94	0.01	0.00
第 2-3 題	0.91	0.02	0.00
第 2-4 題	0.95	0.02	0.00
第 2-5 題	0.87	0.03	0.00
第 2-6 題	0.97	0.01	0.00

表 1-4 流利精通級口語測驗各題因素負荷量摘要表

試題編號	因素負荷量	標準誤	P 值
第 1-1 題	0.84	0.042	0.000
第 2-1 題	0.95	0.022	0.000
第 2-2 題	0.96	0.016	0.000
第 2-3 題	0.94	0.022	0.000
第 3-1 題	0.88	0.039	0.000
第 3-2 題	0.96	0.020	0.000
第 3-3 題	0.91	0.029	0.000

表 1-5 進階高階級寫作測驗單因素模式各向度因素負荷量摘要表

題型-向度名稱	因素負荷量	標準誤	P 值
書信寫作-任務完成度	0.70	0.09	0.00
書信寫作-文意連貫性	0.90	0.05	0.00
書信寫作-形式語言適切性	0.96	0.04	0.00
觀點論述-任務完成度	0.90	0.05	0.00
觀點論述-文意連貫性	0.82	0.04	0.00
觀點論述-形式語言適切性	0.90	0.05	0.00

表 1-6 流利精通級寫作測驗單因素模式各向度因素負荷量摘要表

題型-向度名稱	因素負荷量	標準誤	P 值
摘要寫作-任務完成度	0.95	0.03	0.00
摘要寫作-文意連貫性	0.94	0.03	0.00
摘要寫作-形式語言適切性	0.99	0.02	0.00
觀點論述-任務完成度	0.87	0.06	0.00
觀點論述-文意連貫性	0.92	0.06	0.00
觀點論述-形式語言適切性	0.82	0.05	0.00

附件 2 口語測驗考生自評問卷

入門基礎級口語能力自我評量

1. 我能用簡單的詞彙說出人物和地點，但句子大多零散。I can produce simple mainly isolated phrases about people and places.
 - ☐很少可以 Rarely
 - ☐不常可以 Not often
 - ☐有時可以 Sometimes
 - ☐常常可以 Often
 - ☐總是可以 Always

2. 我能描述我自己、我做的工作和住在什麼地方。I can describe myself, what I do and where I live.
 - ☐很少可以 Rarely
 - ☐不常可以 Not often
 - ☐有時可以 Sometimes
 - ☐常常可以 Often
 - ☐總是可以 Always

3. 我能用簡單的句子談談我的家、我家人和我認識的人。I can talk about my home, my family and people I know in simple sentences.
 - ☐很少可以 Rarely
 - ☐不常可以 Not often
 - ☐有時可以 Sometimes
 - ☐常常可以 Often
 - ☐總是可以 Always

4. 我能描述我平常做什麼事情。I can describe what I do regularly.
 - ☐很少可以 Rarely
 - ☐不常可以 Not often
 - ☐有時可以 Sometimes
 - ☐常常可以 Often
 - ☐總是可以 Always

5. 我能描述我以前做過什麼事情。I can describe what I did in the past.
 - ☐很少可以 Rarely
 - ☐不常可以 Not often
 - ☐有時可以 Sometimes
 - ☐常常可以 Often
 - ☐總是可以 Always

6. 我能描述我打算要做的事情和計畫。I can describe plans, arrangements and alternatives.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
7. 我能大致說出我已經做過的事情。I can tell the main points about something I have done.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
8. 我能簡短地描述一個簡單的故事或一件事情。I can give short simple descriptions of events or tell a simple story.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
9. 我能簡單地描述事物而且做出比較。I can give simple descriptions of things and make comparisons.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
10. 我能描述參加過的活動和我自己的經驗。(例如：上個週末、上個假日)。I can describe past activities and personal experiences (e.g. the last weekend, my last holiday).
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always

11. 我能描述我的教育背景、我現在或最近的工作。I can describe my educational background, my present or most recent job.

- 很少可以 Rarely
- 不常可以 Not often
- 有時可以 Sometimes
- 常常可以 Often
- 總是可以 Always

12. 我能談談我曾經上過的學校、我最近學校的情況或我最近的工作。I can talk about which schools I've gone to, my present educational situation or my present job.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
13. 對於一件事情，我能解釋我喜歡什麼和不喜歡什麼。I can explain what I like and don't like about something.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always

進階高階級口語能力自我評量

1. 對於學業或專業領域的話題，我能不費力做出前後連貫的描述。I can give a reasonably fluent description of a subject within my academic or professional field, presenting it as a linear sequence of points.
 - 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
2. 對於我感興趣的或學習相關的領域，我能直接的描述我熟悉的話題。I can give straightforward descriptions on a variety of familiar subjects related to my own fields of interest or study.
 - 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
3. 我能完整的描述我的經驗、感覺和反應。I can talk in detail about my experiences, feelings and reactions.
 - 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
4. 我能描述一本書或一部電影的情節，並說出我的看法。I can talk about the plot of a book or film and give my opinion.
 - 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
5. 我能描述夢想、希望和抱負。I can describe dreams, hopes and ambitions.
 - 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always

6. 我能敘述一個故事。I can tell a story.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
7. 大部分的情況下，我提出的論點容易被理解。I can develop an argument well enough to be followed without difficulty most of the time.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
8. 我能簡短的解釋和說明我的意見和計畫。I can briefly explain and give reasons for my opinions and plans.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
9. 我能提出詳細的理由以支持自己的論點。I can argue for my point of view in detail.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
10. 我能就一個問題提出自己的觀點，並說明不同選擇之下的優點和缺點。I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always

11. 即使聽眾提出了我事前沒準備到的問題，我也能自然地回答。I can depart spontaneously from a prepared text and follow up points raised by an audience.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
12. 我能發展清楚、有邏輯的論點，並使用適當的例子來延伸及支持自己的論點。I can develop a clear coherent argument, linking ideas logically and expanding and supporting my points with appropriate examples.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always
13. 我能對自己有興趣的相當廣泛的主題，清楚、詳細的描述，並使用適當的說明及例子來擴展和支持自己的論點。I can give clear detailed descriptions on a wide range of subjects relating to my field, expanding and supporting ideas with subsidiary points and relevant examples.
- 很少可以 Rarely
 - 不常可以 Not often
 - 有時可以 Sometimes
 - 常常可以 Often
 - 總是可以 Always

流利精通級口語能力自我評量

1. 無論是專業領域，或是具爭議性的議題，我都能詳細地說明自己的看法，並提出評論或建議。
 - ☐很少可以
 - ☐不常可以
 - ☐有時可以
 - ☐常常可以
 - ☐總是可以
2. 在正式場合中，我能提出一份主題複雜的報告，透過清晰、有組織地說明及適當的例子，準確地表達自己的觀點。
 - ☐很少可以
 - ☐不常可以
 - ☐有時可以
 - ☐常常可以
 - ☐總是可以
3. 不論對方使用什麼說話方式，我都能彈性地調整詞彙、語調，以符合當時的情境。
 - ☐很少可以
 - ☐不常可以
 - ☐有時可以
 - ☐常常可以
 - ☐總是可以
4. 討論抽象、複雜的議題時，我能輕鬆地參與討論，並精確、流利地說明自己的觀點和反對其他立場的理由。
 - ☐很少可以
 - ☐不常可以
 - ☐有時可以
 - ☐常常可以
 - ☐總是可以
5. 針對跟我立場不同的看法或批評，我能透過說明和評論的方式，輕鬆且正確地回應或辯論，並說服對方接受自己的立場。
 - ☐很少可以
 - ☐不常可以
 - ☐有時可以
 - ☐常常可以
 - ☐總是可以
6. 遇到抽象、複雜或不熟悉的主題時，我能整合目前已知的資訊，做出清晰、有組織的報告，強調重點及相關細節，幫助聽眾注意重點。
 - ☐很少可以

- ☐不常可以
- ☐有時可以
- ☐常常可以
- ☐總是可以

7. 針對相關主題，我能整合不同來源的資料和論點，進行全面性概述，總結意見，並以適當的方式結束發言。

- ☐很少可以
- ☐不常可以
- ☐有時可以
- ☐常常可以
- ☐總是可以

8. 在正式場合中發表言論時，我能靈活、精確地運用成語。

- ☐很少可以
- ☐不常可以
- ☐有時可以
- ☐常常可以
- ☐總是可以

附件 3 寫作測驗流利精通級問卷

流利精通級寫作能力自我評量

1. 我能運用複雜的句型寫文章。
 - ☐能做到
 - ☐大概能做到
 - ☐不太能做到
 - ☐做不到

2. 我能運用書面語寫文章。
 - ☐能做到
 - ☐大概能做到
 - ☐不太能做到
 - ☐做不到

3. 在書寫摘要時，我能重新組織長篇文本的資料。
 - ☐能做到
 - ☐大概能做到
 - ☐不太能做到
 - ☐做不到

4. 在書寫摘要時，我能抓住長篇文本資料的重點。
 - ☐能做到
 - ☐大概能做到
 - ☐不太能做到
 - ☐做不到

5. 在論述觀點時，我能對特定議題提出個人看法，並做出結論。
 - ☐能做到
 - ☐大概能做到
 - ☐不太能做到
 - ☐做不到

6. 在論述觀點時，我能對特定議題提出解決方法。
 - ☐能做到
 - ☐大概能做到
 - ☐不太能做到
 - ☐做不到

附件 4 評分會議評分人員調查問卷

表 4-1 進階高階級口語測驗評分人員問卷調查結

問卷題目	描述題		說明題	
	平均數	同意百分比	平均數	同意百分比
1. 評分會議中，會議帶領者對於音檔試評及試評結果討論的形式，說明得很清楚。	5.0	100.0%	5.0	100.0%
2. 評分會議中，會議帶領者對於評分原則、標準音檔及能力指標的內容解說清楚。	5.0	100.0%	5.0	100.0%
3. 承上，評分原則、標準音檔及能力指標的解說，有助於我了解各等級使用者的口語表現。	5.0	100.0%	5.0	100.0%
4. 評分會議中，會議帶領者對於評分原則、標準音檔及能力指標如何應用在評分過程，說明得很清楚。	5.0	100.0%	5.0	100.0%
5. 承上，評分原則、標準音檔及能力指標應用技巧的解說，有助於我進行評分。	5.0	100.0%	5.0	100.0%
6. 評分會議中，試評音檔並就試評結果進行討論，有助於我掌握評分原則以進行評分。	4.8	96.0%	4.7	94.0%
7. 評分會議中，共同討論細則，有助於我掌握評分原則進行評分。	4.7	94.0%	4.6	92.0%
8. 我對於自己在評分階段所評定的音檔成績及評分依據有信心。	4.0	80.0%	4.0	80.0%

註：問卷填答方式，「1」表示非常不同意，「2」表示不同意，「3」表示普通，「4」表示同意，「5」表示非常同意。

表 4-2 流利精通級口語測驗評分人員問卷調查結果

問卷題目	平均數	同意百分比
1. 評分會議中，會議帶領者對於音檔試評及試評結果討論的形式，說明得很清楚。	5.0	100%
2. 評分會議中，會議帶領者對於評分原則、標準音檔及能力指標的內容解說清楚。	5.0	100%
3. 承上，評分原則、標準音檔及能力指標的解說，有助於我了解各等級使用者的口語表現。	5.0	100%
4. 評分會議中，會議帶領者對於評分原則、標準音檔及能力指標如何應用在評分過程，說明得很清楚。	5.0	100%
5. 承上，評分原則、標準音檔及能力指標應用技巧的解說，有助於我進行評分。	5.0	100%
6. 評分會議中，試評音檔並就試評結果進行討論，有助於我掌握評分原則以進行評分。	5.0	100%
7. 評分會議中，共同討論細則，有助於我掌握評分原則進行評分。	4.5	90.0%
8. 我對於自己在評分階段所評定的音檔成績及評分依據有信心。	4.0	80.0%

註：問卷填答方式，「1」表示非常不同意，「2」表示不同意，「3」表示普通，「4」表示同意，「5」表示非常同意。

表 4-3 進階高階級寫作測驗評分人員問卷調查結果

問卷題目	書信寫作		觀點論述	
	平均數	同意百分比	平均數	同意百分比
1. 我了解此次華測會舉辦【評分會議】的目的。	4.0	100.0%	4.0	100.0%
2. 會議帶領者對於評分原則與細則說明得很清楚、詳盡。	4.0	100.0%	4.0	100.0%
3. 會議帶領者進行練習卷評閱及團體討論，讓我更清楚掌握評分原則與細則，並且有助於正式評閱工作。	4.0	100.0%	4.0	100.0%
4. 我根據評分原則與細則進行正式評閱工作。	3.7	91.7%	4.0	100.0%
5. 我對於自己在此次正式評閱中所評定的成績有信心。	3.7	91.7%	3.7	91.7%

註：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。

表 4-4 流利精通級寫作測驗評分人員問卷調查結果

問卷題目	摘要寫作		觀點論述	
	平均數	同意百分比	平均數	同意百分比
1. 我了解此次華測會舉辦【評分會議】的目的。	4.0	100%	4.0	100%
2. 會議帶領者對於評分原則與細則說明得很清楚、詳盡。	4.0	100%	4.0	100%
3. 會議帶領者進行練習卷評閱及團體討論，讓我更清楚掌握評分原則與細則，並且有助於正式評閱工作。	4.0	100%	4.0	100%
4. 我根據評分原則與細則進行正式評閱工作。	4.0	100%	4.0	100%
5. 我對於自己在此次正式評閱中所評定的成績有信心。	3.3	83.3%	3.5	87.5%

註：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。