華語文能力測驗技術報告—2022（3）

口語測驗、寫作測驗信效度

國家華語測驗推動工作委員會

目錄

表目錄

圖目錄

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(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）。

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自身的立即需求」、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 口寫測驗基本能力描述

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	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。	能寫出簡單、不連貫的短語和句子。
基礎級	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內容。	能運用簡單的連接詞寫出簡短的電子郵件，表達立即的需求，如：感謝、道歉、邀請等。
進階級	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；	能寫出詳細的私人信件，藉由描述經驗、情感、事件等，傳達切身相關的訊息。

	能提出簡短的理由支持自己的看法。	
高階級	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。	能撰寫闡述論點的文章或報告，對特定觀點提出支持或反對的理由，並解釋各種面向的優劣。
流利級	在正式場合中，針對複雜、抽象、不熟悉的話題： 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。	1.能大致完整地撰寫貼近原意的摘要，對各種主題的長篇文本資料，能大致適切地重新組織，並能大致使用多種句型與常用書面語。通篇脈絡大致清楚且文句通暢。 2.能撰寫闡述論點的長篇文章或報告，對各種議題能以清楚的邏輯結構予以評析、提出解決方案與作出結論，以表達個人觀點，並能運用多種不同的句型與常用書面語。
精通級	在正式場合中，針對複雜、抽象、不熟悉的話題： 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。	1.能撰寫表達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，並能自如地運用複雜的句型與多樣的高程度書面語。通篇脈絡清晰且文句簡練。 2.能撰寫闡述論點的長篇文章或報告，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，並能自如地運用多種不同的複雜句型與高程度書面語。

（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	4	無	30秒
	經驗描述	3	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的寫作能力有「短語或單句層次之寫作能力」與「簡單描述段落層次之寫作能力」兩大類，寫作測驗根據前者設計了三種題型，後者則設計了「書信寫作」題型。進階高階級測驗希望測得應試者能撰寫「描述個人經驗的信件、訊息」或「闡述特定觀點」的文章，因此設計了針對書信寫作與表達觀點的兩類題型，流利精通級則希望測得應試者「撰寫長篇文章論述自身的觀點」，以及「重新組織資訊凸顯重點」的能力，因此設計了「摘要寫作」與「觀點論述」。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	句子重組	2	20字以內	共20分鐘
	完成對話	2		
	圖片描述	4		
	書信寫作	1	70-120字	20分鐘
進階高階級	書信寫作	1	250-350字	40分鐘
	觀點論述	1	500-600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。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	8-15分
	基礎級	16-24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均採取分析式評分。以流利精通級的觀點論述和題型為例，評分教師依據評分原則和細則，針對「任務完成度」與「語言表現」兩大向度給分，再得出總分。三等級寫作測驗的各大題，皆在此基礎之上為各題型訂定相應的評分原則。

入門基礎級第一部分「句子重組」、「完成對話」及「圖片描述」三種題型，每一題的級分級距設定為0至3級分，第二部分「書信寫作」每一題的評分級距依照寫作表現的不同設定為0至5級分。進階高階級以及流利精通級各題型的級分級距皆為每題0至5級分。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	26-38分
	基礎級	39-49分
進階高階級	進階級	5-7分
	高階級	8-10分
流利精通級	流利級	4-7分
	精通級	8-10分

二、測驗標準化流程

本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、口寫評分與成績檢核、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」。

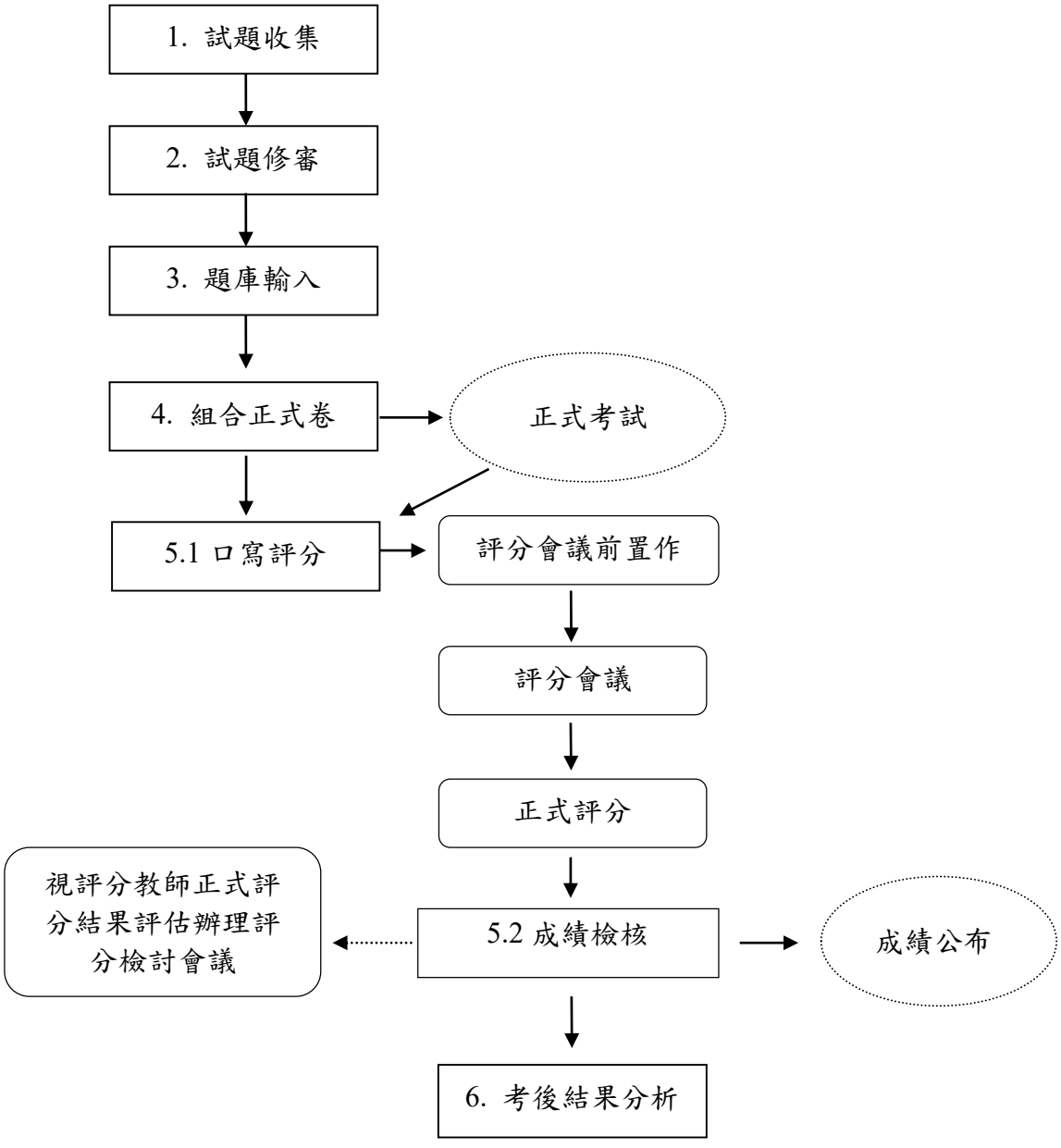

圖 1 測驗標準化流程

口寫測驗「測驗標準化」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. 口寫評分/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：口寫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。
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。

口寫測驗「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。本節以本年度以三月入門基礎級、進階高階級與五月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人員分別有 10 位及 15 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，三個測驗等級評分者嚴格度落在±0.5 logit 以內的比例介於 90%至 100%之間，表示絕大多數評分者嚴格度適中。此外，三個等級評分人員之間的斯皮爾曼等級相關具有高度的一致性，整體看來，評分者間信度良好。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。

寫作測驗入門基礎級、進階高階級與流利精通級皆分為二類題型，各題型分別邀請 3 到 5 位評分人員參與閱卷工作[1]。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嚴格度落在±0.5 logit 以內的比例均為 100%，顯示所有評分者細項給分的嚴格度適中；流利精通級兩個題型介於±0.5 logit 之間的比例為 75%及 60%。斯皮爾曼等級相關結果，進階高階級兩個題型和流利精通級觀點論述題型整體級分相關係數平均值接近或高於.7，流利精通級摘要寫作題型評分者間相關係數平均值未能達到高度相關。整體來說，大部分題型評分者間信度大致良好，少數題型信度表現尚可。在評分者內一致性方面，所有評分人

[1] 由於寫作測驗單句表達多有明確答案，故由 2 名研發人員自行評分，未召開評分會議。

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。同時也會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。此外，華測會每年針對正式評分教師辦理進階評分培訓，以持續強化評分教師對於評分原則掌握度。

（一）評分者間信度（嚴格度分析）

以軟體 Facets[2]部分給分模式（Partial Credit Model，以下簡稱 PCM）[3]對資料進行多面向 Rasch 分析（many facet Rasch measurement，以下簡稱 MFRM）[4]。

口語測驗方面，入門基礎級、進階高階級、流利精通級評分的嚴格度結果，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 10 位評分者嚴格度均介於±0.5 logit 以內，表示所有評分者嚴格度相近，比例達 100%；5 月場次 15 位評分者嚴格度除 A32 給分較為嚴格外，其餘評分者嚴格度落在±0.5 logit 以內，評分嚴格度相近的比例為 93.3%。進階高階級口語測驗方面，B50 給分較為寬鬆，其餘 9 位評分者嚴格度均介於±0.5 logit 以內，比例為 90%。流利精通級測驗 4 位評分者的嚴格度介於±0.5 logit 以內，表示所有評分者嚴格度相近。整體而言，三個等級測驗的評分者間信度良好。

表 8　3 月入門基礎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
A26	80	1.80	1.83	0.157	0.204	0.87
A22	80	1.80	1.84	0.115	0.205	0.86
A05	80	1.80	1.84	0.115	0.205	1.00
A16	80	1.80	1.85	0.073	0.205	0.84
A18	80	1.80	1.86	0.031	0.206	1.00
A20	80	1.90	1.88	-0.012	0.206	0.88

[2] Linacre, J. M. (2013). Facets® (Version 3.71.3)[Computer Software]. Beaverton, Oregon: Winsteps.com.
[3] Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
[4] Linacre, J. M. (1988). *Many-facet Rasch measurement*. Chicago: MESA Press.

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
A32	80	1.90	1.88	-0.012	0.206	0.98
A02	80	1.90	1.89	-0.054	0.206	0.80
A31	80	1.90	1.90	-0.097	0.207	0.93
A25	80	1.90	1.96	-0.314	0.210	0.87

表 9　5 月入門基礎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
A32	288	1.00	1.05	0.656	0.098	0.74
A31	288	1.20	1.18	0.415	0.091	0.80
A22	288	1.30	1.34	0.142	0.087	0.97
A04	288	1.30	1.34	0.142	0.087	0.92
A23	288	1.40	1.45	-0.050	0.086	0.69
A24	288	1.60	1.45	-0.056	0.088	0.91
A26	288	1.40	1.46	-0.065	0.086	0.72
A18	288	1.30	1.46	-0.065	0.091	1.13
A20	288	1.40	1.46	-0.069	0.089	0.83
A14	32	1.00	1.48	-0.110	0.324	0.74
A16	755	1.60	1.46	-0.110	0.056	0.73
A02	288	1.20	1.50	-0.145	0.090	0.95
A25	288	1.60	1.51	-0.173	0.089	0.89
A27	288	1.30	1.53	-0.210	0.090	0.94
A05	288	1.70	1.58	-0.299	0.089	0.81

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
B04	162	2.90	2.72	0.436	0.131	0.70
B18	156	2.80	2.79	0.285	0.132	0.83
B48	156	2.80	2.81	0.233	0.132	0.73
B49	156	3.00	2.83	0.198	0.130	1.11
B39	156	3.00	2.84	0.181	0.130	0.65
B07	126	3.10	2.84	0.164	0.148	0.94
B43	162	3.00	2.92	0.006	0.132	0.68
B16	240	3.00	2.98	-0.128	0.105	0.60
B47	156	2.90	2.98	-0.135	0.132	0.71
B50	156	3.50	3.48	-1.239	0.135	1.25

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
C07	273	2.00	1.83	0.304	0.097	0.96
C06	329	2.00	1.92	0.078	0.095	0.76
C02	273	2.20	2.02	-0.185	0.095	1.29
C05	57	0.30	2.04	-0.196	0.614	0.85

寫作測驗方面，評分者在入門基礎級、進階高階級、流利精通級各題型細項分數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，入門基礎級單句表達 2 名研發人員、書信寫作 4 名評分者嚴格度均介於±0.5 logit 之間，顯示所有評分者嚴格度相近，比例達 100%。進階高階級測驗，書信寫作與觀點論述 5 名與 3 名評分者的嚴格度分布同樣落在±0.5 logit 範圍以內，比例均達 100%，表示所有評分者嚴格度接近。此二個測驗等級各題型的評分者間信度良好。流利精通級測驗，摘要寫作和觀點論述 4 名與 5 名評分者嚴格度落在±0.5 logit 範圍以內的百分比為 75%以及 60%，評分者間信度尚可。其中，摘要寫作題型 C07 給分較為寬鬆，而觀點論述題型 C14 較為嚴格、C03 較為寬鬆。

表 12 入門基礎級寫作測驗單句表達評分者嚴格度

評分者	評閱題數	觀察的平均值	調整後平均值	嚴格度	標準誤	INFIT MNSQ
A36	80	2.6	2.58	0.083	0.287	1.02
A27	80	2.6	2.62	-0.083	0.291	0.98

表 13 入門基礎級寫作測驗書信寫作評分者嚴格度

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
A36	8	3.60	3.73	0.221	0.228	0.70
A14	8	3.60	3.75	0.169	0.229	0.96
A40	8	3.70	3.80	0.008	0.234	0.64
A37	8	3.80	3.92	-0.397	0.248	1.34

表 14 進階高階級寫作測驗書信寫作評分者嚴格度

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
B20	29	2.8	2.85	0.310	0.087	0.82
B43	29	2.8	2.93	0.203	0.088	1.26
B49	29	2.9	3.08	-0.015	0.089	0.78
B39	29	3.0	3.20	-0.191	0.090	0.85
B25	29	3.1	3.27	-0.306	0.091	1.05

表 15 進階高階級寫作測驗觀點論述評分者嚴格度

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
B14	29	3.0	3.1	0.099	0.090	0.94
B31	29	3.1	3.13	0.058	0.091	0.95
B07	29	3.2	3.29	-0.158	0.092	1.26

表 16 流利精通級寫作測驗摘要寫作評分者嚴格度

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	8	1.70	2.05	0.398	0.259	0.74
C21	48	2.00	2.06	0.337	0.096	0.97
C23	40	2.10	2.10	0.200	0.103	1.04
C07	48	2.30	2.39	-0.934	0.099	1.07

表 17 流利精通級寫作測驗觀點論述評分者嚴格度

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
C14	30	1.6	1.65	0.635	0.091	0.63
C04	30	1.8	1.86	0.177	0.087	0.93
C24	50	2.0	2.01	-0.120	0.064	0.80
C25	30	2.1	2.02	-0.143	0.078	1.43
C03	30	2.3	2.22	-0.549	0.078	1.22

（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 0.4 以上表示有中度相關，達 0.7 以上則表示有高度相關。入門基礎級六組評分者相關係數整體平均值達.7 或.8 以上，具有高度正相關；進階高階級三個組別的表現相仿，整體平均值都達到.8 或.9 以上，具高度正相關；流利精通級未分組，評分者相關係數整體平均值為.913。

整體來看，入門基礎級、進階高階級與流利精通級測驗，兩兩評分者之間給分有高度正相關存在，評分者間信度良好。且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整體平均值

測驗等級	第 1 組	第 2 組	第 3 組	第 4 組	第 5 組	第 6 組
入門基礎級	0.788	0.718	0.764	0.754	0.810	0.822
進階高階級	0.902	0.825	0.910	--	--	--
流利精通級	0.913	--	--	--	--	--

註：因 3 月入門基礎級測驗考生人數未達 20 人，未進行分析。

寫作測驗針對進階高階級與流利精通級各題型評分者兩兩進行斯皮爾曼等級相關分析的結果如表 19 所示，因篇幅有限，僅呈現每個題型整體級分評分者間相關係數的平均值。進階高階級書信寫作與觀點論述題型、流利精通級觀點論述題型的整體級分相關係數平均值依序為 0.716、0.699 以及 0.687，接近或具有高度正相關；流利精通級摘要寫作題型評分者整體級分的相關係數平均值為 0.518，具有中度正相關。

整體來看，進階高階級與流利精通級觀點論述題型評分者間信度大致良好；流利精通級摘要寫作評分者間信度則表現尚可，然本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	整體級分相關係數平均值
進階高階級	書信寫作	0.716
	觀點論述	0.699
流利精通級	摘要寫作	0.518
	觀點論述	0.687

註：因入門基礎級測驗考生人數未達 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 月正式考試各評

分者之標準誤介於 0.204 至 0.210，相差不大；5 月正式考試評分者 A14 因為核心評分者，評閱音檔數較少，故標準誤較大，扣除此名評分者，其餘評分者之標準誤介於 0.056 至 0.091 之間，差異不大，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 0.5 至 1.5 之間，兩次考試所有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，各評分者之標準誤介於 0.105 至 0.148，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，C05 同樣為核心評分者，評閱音檔數較少因而標準誤較大，除了 C05 之外的三位評分者標準誤相近，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，入門基礎級單句表達與書信寫作分析結果如表 12、表 13 所示，單句表達二名評分者標準誤分別為 0.287、0.291，書信寫作標準誤介於 0.228 至 0.248 之間，差異不大，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 0.5 至 1.5 之間，亦表示評分者內一致性佳，自身評分穩定性良好。進階高階級二個題型的結果見表 14 及表 15，各個評分者標準誤介於 0.087 至 0.091 之間，以及 0.090 至 0.092 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。至於流利精通級二個題型分析結果如表 16、表 17 所示，C26 由於臨時接替 C23 完成評分工作，評閱篇數較少，故標準誤較大，其餘三位評分者的標準誤介於 0.096 至 0.103 之間，差異不大，觀點論述評分者標準誤介於 0.064 至 0.091 之間；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。

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 0.5 到 1.5 以及 INFIT ZSTD 介於 -3.0 到 3.0 的標準評估試題是否與模式適配，分析結果口語測驗入門基礎級、進階高階級、流利精通級試題適配率為 100%、100%以及 86%，顯示絕大多數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，分別有八項和七項評分細項；流利精通級摘要寫作和觀點論述題型，分別有 11 項和 15 項評分細項（細項名稱參見附件 1），適配率依序為 100%、100%、100%以及 93%，表示絕大多數細項 INFIT MNSQ 以及 INFIT ZSTD 數值皆符合模式，顯示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	8	8	100%
	進階高階級	6	6	100%
	流利精通級	7	6	86%
寫作	進階高階級書信寫作	8	8	100%
	進階高階級觀點論述	7	7	100%
	流利精通級摘要寫作	11	11	100%
	流利精通級觀點論述	15	14	93%

註：3 月入門基礎級口語及寫作測驗因考生人數未達 20 人，未進行分析。

2. 驗證性因素分析

此節使用 Mplus[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。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；寫作測驗各題型同樣為單因素模式，測量寫作表達能力。由於篇幅有限，在此僅呈現入門基礎級口語測驗和進階高階級寫作測驗書信寫作題型結果，其餘測驗等級的正式考試因素負荷量等參數請見附件 1。

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.76 至.88

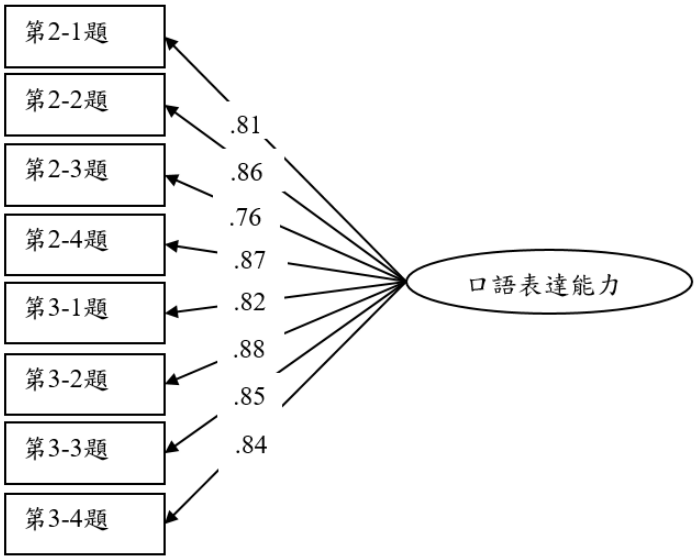

圖 2 入門基礎級口語測驗單因素模式

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書信寫作單因素模式分析結果如圖 3 所示，各評分細項的因素負荷量介於.50 至.94 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.

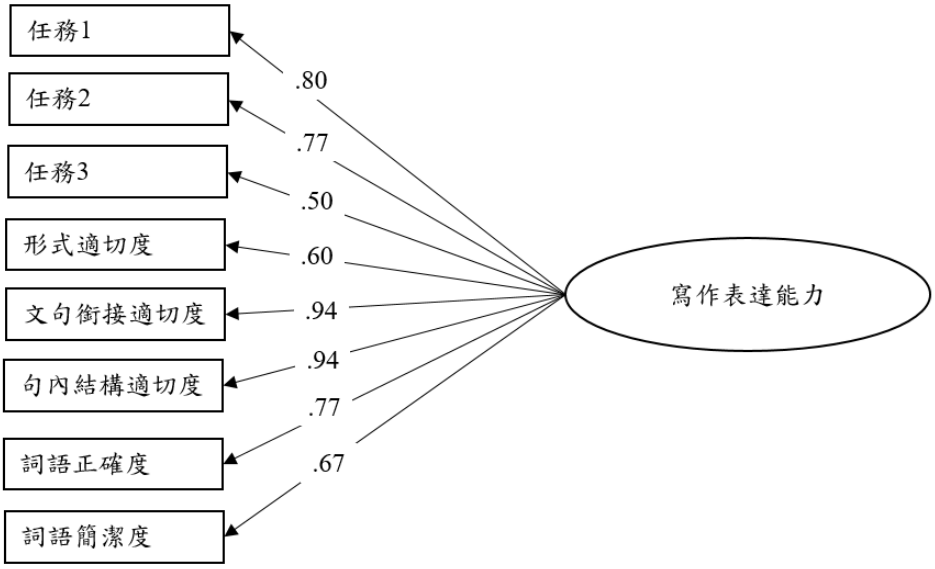

圖 3 進階高階級寫作測驗書信寫作單因素模式

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，由表 21 可知，口語、寫作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。絕對適配度評估標準為平均概似平方誤根係數（root mean square error of approximation，以下簡稱 RMSEA）小於.08，下表雖然多數模式的 RMSEA 都大於.08，但文獻指出，RMSEA 指標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）[6]，而除入門基礎級口語測驗外，其餘各個模式的考生人數均小於 100 人，可能因此導致 RMSEA 數值較高。至於增值適配度評估部分，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。

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，口語、寫作測驗之各等級具有一定的建構效度，各道試題/評分向度反映測得一致之口語和寫

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

作表達能力。

表 21 口語測驗、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
口語	入門基礎級單因素模式	2.87	0.098	0.989	0.984
	進階高階級單因素模式	0.67	0.000	1.000	1.001
	流利精通級單因素模式	1.22	0.070	1.000	0.999
寫作	進階高階級書信寫作單因素模式	1.81	0.165	0.977	0.968
	進階高階級觀點論述單因素模式	1.30	0.100	0.997	0.996
	流利精通級摘要寫作單因素模式	2.30	0.158	0.991	0.989
	流利精通級觀點論述單因素模式	2.05	0.142	0.979	0.976

註：3 月入門基礎級口語及寫作測驗因考生人數未達 20 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，進階高階級、流利精通級考生自評結果與測驗總分的積差相關係數分別為.655 以及.748，均達到.01 顯著水準，顯示各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。而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，進階高階級 17 題的自評問題與測驗總分相關係數介於.360 至.687 之間，都達顯著水準（p<.05 或 p<.01），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，8 題自評問題與測驗總分相關係數介於.613 至.769 之間，均達到顯著水準（p<.01），同樣具有良好的效標關聯效度。詳細問卷內容請見附件 2。

表 22 口語測驗自評問卷各題與測驗總分之相關分析

進階高階級 (N=39)	Q1	Q2	Q3	Q4	Q5	Q6	Q7
	.556**	.649**	.592**	.687**	.553**	.505**	.563**
	Q8	Q9	Q10	Q11	Q12	Q13	Q14
	.366*	.360*	.441**	.496**	.433**	.418**	.610**
	Q15	Q16	Q17				
	.490**	.448**	.576**				
流利精通級 (N=47)	Q1	Q2	Q3	Q4	Q5	Q6	Q7
	.706**	.761**	.686**	.681**	.613**	.616**	.769**
	Q8						
	.687**						

註 1： *p<.05；**p<.01。

註 2：3 月入門基礎級測驗因考生人數未達 30 人，樣本代表性較不足，故未進行分析。

寫作測驗方面，進階高階級、流利精通級考生自評結果與測驗總分的積差相關係數分別為.348 以及.066，均未達到.05 顯著水準，顯示考生自評寫作能力與測驗總分之間沒有正相關存在。而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 23 所示，進階高階級 Q1.「我能寫一封比較詳細的信，告訴別人自己的經驗。」與 Q2.「我能寫一封比較詳細的信，告訴別人自己對一件事情的看法。」與測驗總分相關係數分別為.591(p<.01)，與.431(p<.05)，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；而 Q3.「寫一篇表達自己想法的文章或報告時，我能對某個觀點提出支持或反對的理由。」及 Q4.「寫一篇表達自己想法的文章或報告時，我能適當地強調重點和部分細節，而且清楚地說明原因。」相關係數較低。由於進階高階級測驗即將於明年（112 年）辦理改版後的正式考試，將參考調整後之題型及上述分析結果討論並修改問卷內容。流利精通級部分，6 題自評問題與測驗總分相關係數均未達到顯著水準（p>.01），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人數較少，同時通過率偏低，表示參與本次測驗的大部分考生寫作能力較低並未達流利級，但未通過之考生，在各題均以回答「大概能做到」占多數（比例介於 54.1%至 67.6%之間），其次為「能做到」，僅極少數回答「不太能做到」，而此分布趨勢與通過流利級者相近，以至於相關係數偏低。研發人員與測驗專家討論後推測可能因問題提問方式讓考生不容易理解，將

修改問卷內容，讓考生在易於理解問卷內容的前提下，能更確實地依照自身真實的語言能力表現回答問卷。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

進階高階級	Q1	Q2	Q3	Q4		
(N=30)	.591**	.431*	0.035	0.049		
流利精通級	Q1	Q2	Q3	Q4	Q5	Q6
(N=52)	.120	.188	.001	.088	-.018	-.049

註 1： *p<.05；**p<.01。

註 2：入門基礎級測驗因考生人數未達 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 111 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

在評分的作業流程上，為確保評分品質，口語測驗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.評分流程。 2.分人員共擬評分細則。 3.試評，討論評分結果，建立評分共識。
	正式評分	1.為期二週的評分工作。 2.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 題的同意百分比為 82%,其餘各題同意百分比高達 100%。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	平均數	同意者百分比
1. 評分會議中，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	100%
2. 評分會議中，會議帶領者對於評分原則、標準音檔及相關口語能力指標的內容解說清楚。	5	100%
3. 承上，評分原則、標準音檔及相關口語能力指標的解說，有助於我了解各等級口語使用者的能力表現。	5	100%
4. 評分會議中，會議帶領者對於評分原則、標準音檔及相關口語能力指標在評分過程上的應用技巧說明得很清楚。	5	100%
5. 承上,評分原則、標準音檔及相關口語能力指標應用技巧的解說，有助於我進行正式給分。	5	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	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。	5	100%
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。	4.1	82%

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	同意者百分比
1. 我了解此次華測會舉辦【評分會議】的目的。	4	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	100.0%

問卷題目	平均數	同意者百分比
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	100.0%
4. 我根據評分原則與細則進行正式評閱工作。	4	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	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，華語文口寫測驗之能力描述、測驗題型題數、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信度與效度之測驗效能評估。

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寫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「評分者自身給分穩定性」與「評分者間給分一致性」二個面向評估測驗信度。針對評分嚴格度變異較大以及自身給分一致性不穩定的評分者，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，即不予續聘或改聘為其他適合等級的評分者。此外，也進行評分者偏誤研究，以進一步提升評分結果之一致性。

在測驗效度分析部分，為使評分者皆能遵守測驗所擬定之評分原則，並據此給予受測者適切的評分，本測驗採取了標準化的評分流程來培訓評分者。此標準化流程為程序效度，可確保測驗相關內容皆經由標準化程序而來，為本測驗提供內容效度方面的證據。除了具備測驗之內容效度方面的證據之外，在施測完成後，本會也針對測驗所得之受測者作答反應資料，分別進行了試題分析與驗證性因素

分析，主要目的在於確認受測者之反應資料所建構出的測驗架構，是否與口寫測驗研發之初所制訂的目標相同，並以此作為測驗之建構效度證據。最後，我們還透過受測者自評結果與受測者實際測驗結果的對照，來評估測驗結果的預測力，可以說，口語測驗具有測驗之效標效度證據，寫作測驗本年度效標關聯效度結果不理想，將針對問卷內容進行調整，持續收集效度證據。

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 華語文口語測驗技術報告。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 華語文寫作測驗技術報告。新北市：國家華語測驗推動工作委員會。

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

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

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 口語測驗進階高階級測驗各題因素負荷量及殘差變異量摘要表

試題編號	因素負荷量	標準誤	P 值
第 2-1 題	0.89	0.038	0.000
第 2-2 題	0.92	0.032	0.000
第 2-3 題	0.98	0.021	0.000
第 2-4 題	0.98	0.016	0.000
第 2-5 題	1.01	0.021	0.000
第 2-6 題	0.95	0.028	0.000

表 2 口語測驗流利精通級測驗各題因素負荷量及殘差變異量摘要表

試題編號	因素負荷量	標準誤	P 值
第 1-1 題	0.96	0.016	0.000
第 2-1 題	0.97	0.018	0.000
第 2-2 題	0.98	0.020	0.000
第 2-3 題	0.97	0.022	0.000
第 3-1 題	0.88	0.036	0.000
第 3-2 題	0.96	0.011	0.000
第 3-3 題	0.97	0.016	0.000

表 3 寫作測驗進階高階級測驗觀點論述各細項因素負荷量及殘差變異量摘要表

細項名稱	因素負荷量	標準誤	P 值
起合	0.86	0.057	0.000
脈絡	0.96	0.039	0.000
論證	0.75	0.086	0.000
格式分段	0.56	0.101	0.000
標點	0.62	0.105	0.000
句型句法	0.99	0.025	0.000
詞語	0.96	0.020	0.000

表 4 寫作測驗流利精通級測驗摘要寫作各細項因素負荷量及殘差變異量摘要表

細項名稱	因素負荷量	標準誤	P 值
任務完成度 1_訊息	0.68	0.027	0.000
任務完成度 2_組織	0.98	0.023	0.000
任務完成度 3_不合規定	0.73	0.036	0.000
句型詞藻表現力 1_複雜句型	0.92	0.020	0.000
句型詞藻表現力 2_多樣句型	0.97	0.014	0.000
句型詞藻表現力 3_詞語轉換	0.79	0.031	0.000
句型詞藻表現力 4_冗贅重複	0.99	0.026	0.000
詞彙語法正確度 1_詞彙語法錯誤	0.78	0.038	0.000
詞彙語法正確度 2_增字漏字錯別字	1.00	0.014	0.000
詞彙語法正確度 3_標點符號不當	0.91	0.025	0.000
詞彙語法正確度 4_分行分段不當	0.94	0.019	0.000

表 5 寫作測驗流利精通級測驗觀點論述各細項因素負荷量及殘差變異量摘要表

細項名稱	因素負荷量	標準誤	P 值
任務完成度 1A_開頭	0.76	0.057	0.000
任務完成度 1B_結尾	0.70	0.054	0.000
任務完成度 2A_任務 1 舉證	0.76	0.049	0.000
任務完成度 2B_任務 1 組織	0.85	0.043	0.000
任務完成度 3A_任務 2 措施	0.64	0.061	0.000
任務完成度 3B_任務 2 組織	0.68	0.061	0.000
句型詞藻表現力 1_複雜句型	0.95	0.022	0.000
句型詞藻表現力 2_多樣句型	0.97	0.021	0.000
句型詞藻表現力 3_詞語轉換	0.93	0.023	0.000
句型詞藻表現力 4_冗贅重複	0.79	0.056	0.000
詞彙語法正確度 1_詞彙語法錯誤	0.85	0.041	0.000
詞彙語法正確度 2_增字漏字錯別字	1.01	0.013	0.000
段落形式適切度 1_分段	0.59	0.075	0.000
段落形式適切度 2_空格	1.03	0.044	0.000
段落形式適切度 3_標點	0.77	0.049	0.000

附件 2 口語測驗考生自評問卷

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. 在預先準備的情況下，我能簡短的發表跟我日常生活領域有關的事項。I can deliver short rehearsed announcements and statements on everyday matters within my field.
   ◯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. 在預先準備的情況下，對於我熟悉的主題，我能做出簡單、清楚、易懂的口語。I can give a simple, prepared presentation on a familiar topic within my field that is clear and precise enough to be followed without difficulty most of the time and in which the main points can be understood.
    ◯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

12. 我能建立一個合理的、有邏輯的論點。I can construct a chain of reasoned argument, linking my ideas logically.
    ◯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. 我能做出清晰、有組織的敘述，並能強調重點及相關細節。I can give a clear, systematically developed presentation, with highlighting of significant points and relevant supporting 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

14. 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

15. 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

16. 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

17. 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 寫作測驗進階高階級問卷

進階高階級寫作能力自我評量

1. 我能寫一封比較詳細的信，告訴別人自己的經驗。I am able to write a letter with relative details to tell someone my experience.
   ◯能做到 Yes
   ◯大概能做到 Likely
   ◯不太能做到 Unlikely
   ◯做不到 No

2. 我能寫一封比較詳細的信，告訴別人自己對一件事情的看法。I am able to write a letter with relative details to tell someone my opinion about something.
   ◯能做到 Yes
   ◯大概能做到 Likely
   ◯不太能做到 Unlikely
   ◯做不到 No

3. 寫一篇表達自己想法的文章或報告時，我能對某個觀點提出支持或反對的理由。When writing an essay or report expressing my own ideas, I am able to put forward arguments for or against a particular viewpoint.
   ◯能做到 Yes
   ◯大概能做到 Likely
   ◯不太能做到 Unlikely
   ◯做不到 No

4. 寫一篇表達自己想法的文章或報告時，我能適當地強調重點和部分細節，而且清楚地說明原因。When writing an essay or report expressing my own ideas, I am able to adequately emphasize important points and some details, and clearly explain my reasons.
   ◯能做到 Yes
   ◯大概能做到 Likely
   ◯不太能做到 Unlikely
   ◯做不到 No

流利精通級寫作能力自我評量

1. 書寫摘要時，我能將長篇文本資料適切地重新組織。
   ◯能做到
   ◯大概能做到
   ◯不太能做到
   ◯做不到

2. 書寫摘要時，我能表達資料的原意且突顯重點。
   ◯能做到
   ◯大概能做到
   ◯不太能做到
   ◯做不到

3. 論述觀點時，我能以縝密的邏輯結構，對議題提出個人看法，並作出適當的結論。
   ◯能做到
   ◯大概能做到
   ◯不太能做到
   ◯做不到

4. 論述觀點時，我能對議題提出完善的解決方法。
   ◯能做到
   ◯大概能做到
   ◯不太能做到
   ◯做不到

5. 書寫文章時，我能靈活地運用複雜的句型。
   ◯能做到
   ◯大概能做到
   ◯不太能做到
   ◯做不到

6. 書寫文章時，我能靈活地運用多樣的高程度書面語。
   ◯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 1　進階高階級口語測驗評分人員問卷調查結果

問卷題目	平均數	同意者百分比
1. 評分會議中，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	100%
2. 評分會議中，會議帶領者對於評分原則、標準音檔及相關口語能力指標的內容解說清楚。	5	100%
3. 承上，評分原則、標準音檔及相關口語能力指標的解說，有助於我了解各等級口語使用者的能力表現。	5	100%
4. 評分會議中，會議帶領者對於評分原則、標準音檔及相關口語能力指標在評分過程上的應用技巧說明得很清楚。	5	100%
5. 承上，評分原則、標準音檔及相關口語能力指標應用技巧的解說，有助於我進行正式給分。	5	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	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。	5	100%
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。	4.1	82%

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 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	同意者百分比
1. 評分會議中，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	100%
2. 評分會議中，會議帶領者對於評分原則、標準音檔及相關口語能力指標的內容解說清楚。	5	100%
3. 承上，評分原則、標準音檔及相關口語能力指標的解說，有助於我了解各等級口語使用者的能力表現。	5	100%
4. 評分會議中，會議帶領者對於評分原則、標準音檔及相關口語能力指標在評分過程上的應用技巧說明得很清楚。	5	100%
5. 承上，評分原則、標準音檔及相關口語能力指標應用技巧的解說，有助於我進行正式給分。	5	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	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。	5	100%
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。	4.5	90%

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 3　進階高階級寫作測驗書信寫作題型評分人員問卷調查結果

問卷題目	平均數	同意者百分比
1. 我了解此次華測會舉辦【評分會議】的目的。	4	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	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	100%
4. 我根據評分原則與細則進行正式評閱工作。	4	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	75%

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　進階高階級寫作測驗觀點論述題型評分人員問卷調查結果

問卷題目	平均數	同意者百分比
6. 我了解此次華測會舉辦【評分會議】的目的。	4	100%
7. 會議帶領者對於評分原則與細則說明得很清楚、詳盡。	4	100%
8. 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	100%
9. 我根據評分原則與細則進行正式評閱工作。	4	100%
10.我對於自己在此次正式評閱中所評定的成績有信心。	3	75%

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 5　流利精通級寫作測驗摘要寫作題型評分人員問卷調查結果

問卷題目	平均數	同意者百分比
1. 我了解此次華測會舉辦【評分會議】的目的。	4	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	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	100%
4. 我根據評分原則與細則進行正式評閱工作。	4	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	84%

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 6　流利精通級寫作測驗觀點論述題型評分人員問卷調查結果

問卷題目	平均數	同意者百分比
1. 我了解此次華測會舉辦【評分會議】的目的。	4	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	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	100%
4. 我根據評分原則與細則進行正式評閱工作。	4	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	75%

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。