

# CEFR 與 ACTFL 大綱口語指標對應研究： 以「華語文口語測驗」為媒介\*

廖才儀 藍珮君 梁綺容<sup>1</sup>  
國家華語測驗推動工作委員會

陳柏熹

國立臺灣師範大學教育心理與輔導學系、華語文與科技研究中心

## 摘要

本研究主旨為建立歐洲共同語文參考架構 (CEFR) 與美國外語教學協會能力指標 (ACTFL Proficiency Guidelines) 這兩大架構之口語指標的對應關係，所採用之華語文口語測驗 (TOCFL Speaking test) 是以 CEFR 為基礎；作法為邀集 17 位專家透過 Exemplar Response 法標準設定程序，使用該測驗各題型各級分之標準音檔，連結 CEFR 與 ACTFL 大綱的各等級口語指標，並提出程序性效度、內部效度做為證據。研究結果顯示，CEFR 框架 A1 程度的華語口語能力介於 ACTFL 大綱初級高等至中級初等之間；A2 位於中級初等與中級中等之間；B1 略低於中級高等；B2 落在高級初等與高級中等之間；C1 介於高級高等至優秀級之間；C2 程度高於優秀級。此外，亦發現數項共同特徵可作為未來研究者進行兩個指標對應研究時的參考依據。

**關鍵詞：**ACTFL 大綱 CEFR Exemplar Response 法 口語測驗  
標準設定

---

\* 本研究感謝教育部補助「107 年度華語文能力測驗」計畫，感謝協助本研究進行標準設定會議所有專家學者提供之寶貴意見，以及本刊兩位匿名審查委員給予之寶貴建議，特此致謝。

<sup>1</sup> 本文通訊作者。

## 1. 緒論

國際上評估外語能力的主要框架有兩個：美國外語教學協會能力指標 (ACTFL Proficiency Guidelines, 簡稱 ACTFL 大綱), 以及歐洲共同語文參考架構 (The Common European Framework of Reference for Languages, 簡稱 CEFR), 兩個框架都以溝通任務做為能力描述的基礎, 也都以提供共通的語言能力描述為研製目的, 已共存約二十年, 雖制訂方向及理論相近, 但制訂機構對等級定義、各項能力分級描述則有所差異, 彼此間無法直接對應, 需透過實證研究成果才能明確這兩套語言指標間的對應關係。隨著跨地域學習的風潮, 為了便於來自不同背景、區域, 但都以同一種目標語做為學習、教學或測驗研發對象者釐清各語言測驗與 CEFR、ACTFL 大綱的對應關係, 進而滿足語言課程安排或了解測驗結果具體意義等實用需求, 近十年才陸續有學者進行相關研究, 如 Waldemar (2010)、Bärenfänger 和 Tschirner (2012)、American Council on the Teaching of Foreign Languages (2016)、Language Testing International (2016)、趙家璧、藍珮君和陳柏熹 (2016) 等等。

華語文口語測驗 (TOCFL Speaking test) 為國家華語測驗推動工作委員會 (簡稱華測會) 研發的標準化語言測驗, 主要參酌 CEFR; 此測驗將華語口語能力分為四等八級: 四等由低至高依序為準備級 (Band Novice)、入門基礎級 (Band A)、進階高階級 (Band B) 及流利精通級 (Band C); 每一等再依測驗成績細分兩級, 依序為準備級一級 (Novice 1)、準備級二級 (Novice 2)、入門級 (Level 1)、基礎級 (Level 2)、進階級 (Level 3)、高階級 (Level 4)、流利級 (Level 5) 及精通級 (Level 6), 共八級。入門基礎級至流利精通級參酌自 CEFR, 並在 2013 至 2016 年透過標準設定程序訂定測驗通過門檻, 確認可分別對應至 CEFR 之 A1 (Breakthrough)、A2 (Waystage)、B1 (Threshold)、B2 (Vantage)、C1 (Effective operational proficiency) 與 C2 (Mastery) (國家華語測驗推動工作委員會 2015, 2016, 2018); 準備級則是在 2017 年初因應海外地區零起點學習者需求所擴增的等級, 但因當時 CEFR 尚未公布更低等級之指標, 故準備級指標之內容參酌自 ACTFL 大綱, 並透過標準設定程序確認準備級一級、準備級二級分別可對應至 ACTFL 大綱之 Novice Low 與 Novice Mid (國家華語測驗推動工作委員會 2017)。

為促使華語文口語測驗已對應至 CEFR 之 A1 至 C2 等級的入門基礎級至流利精通級架構也能與 ACTFL 大綱之框架接軌, 有利於熟悉 ACTFL 大綱的受測者也能了解該測驗成績的代表意義; 故本研究以該測驗各等級標準音檔

為研究工具，透過量化方式建立華語文口語測驗與 ACTFL 大綱之對應關係；同時，也透過比對兩者能力描述內容之方式，確認 CEFR 與 ACTFL 大綱在口語表達能力、語言表現、各項溝通任務掌握進程等方面是否具有共同特徵，進而為 CEFR 與 ACTFL 大綱之對應關係取得更多參考依據。

## 2. 文獻探討

本節將先概述 CEFR、ACTFL 大綱之內容，並探究語言測驗與語言框架對應、標準設定程序等相關文獻，以尋求本研究可參考採用之方法。

### 2.1 CEFR 與 ACTFL 大綱中的口語指標內容

#### 2.1.1 CEFR 各等級口語指標

CEFR 係由歐洲理事會（Council of Europe，簡稱 CoE）編制，目的為替教育體系與課程安排迥異的歐洲各國建立一套跨語言間互相認可的共通標準。最初公布於 2001 年，2018 年公布增修版。

2001 年的版本將外語學習者的語言能力分為三等六級，三等依序為「初級使用者」(Basic User)、「獨立使用者」(Independent User) 與「精通使用者」(Proficient User)；每等之下各自細分兩級為 A1 (Breakthrough)、A2 (Waystage)、B1 (Threshold)、B2 (Vantage)、C1 (Effective operational proficiency) 及 C2 (Mastery)；語言溝通能力則被分為聽、讀、口、寫四類。

CEFR 在此框架之下，針對外語學習者和使用者的口語能力表現，制訂出一套含括各等級口語表達、口語互動及溝通策略等語言能力的系統性總表 (Council of Europe 2001)。2018 年的增修版中，除了更新或擴充部分語言活動項目的能力進展描述之外，也向下擴增了 pre-A1 等級，使 CEFR 成為三等七級的架構 (Council of Europe 2018)。

#### 2.1.2 ACTFL 大綱各等級口語指標

ACTFL 大綱為美國普遍採用之外語框架，係由美國外語教師協會 (American Council on the Teaching of Foreign Languages, ACTFL) 改寫美國跨部門語言圓桌量表 (Interagency Language Roundtable scale, ILR) 而成；首次公布於 1986 年，旨在提供全球一套可用於學術和工作場合評估語言溝通能力的標準；後經 1999、2001 年兩次修訂，現行版本則為 2012 年增修版。

其架構為五等十一級：五個主要等級由低至高依序為初級 (Novice，簡

稱 N)、中級 (Intermediate, 簡稱 I)、高級 (Advanced, 簡稱 A)、優秀級 (Superior, 簡稱 S) 和優異級 (Distinguished, 簡稱 D), 各自代表一個具體的能力範圍, 也都涵蓋低於它的所有等級; N、I、A 這三個主要等級之下, 又再各自細分了初等 (Low)、中等 (Mid) 與高等 (High) 三個次要等級, 如圖 1 所示, 五等十一級名稱依序為初級初等 (Novice Low, NL)、初級中等 (Novice Mid, NM)、初級高等 (Novice High, NH)、中級初等 (Intermediate Low, IL)、中級中等 (Intermediate Mid, IM)、中級高等 (Intermediate High, IH)、高級初等 (Advanced Low, AL)、高級中等 (Advanced Mid, AM)、高級高等 (Advanced High, AH)、優秀級 (S)、優異級 (D)。在此框架之下, ACTFL 大綱提供了個人在真實情境中, 聽、讀、口、寫這四項語言能力在各等級中可以熟練達到的程度, 以及還無法達到的侷限性 (American Council on the Teaching of Foreign Languages 2012)。

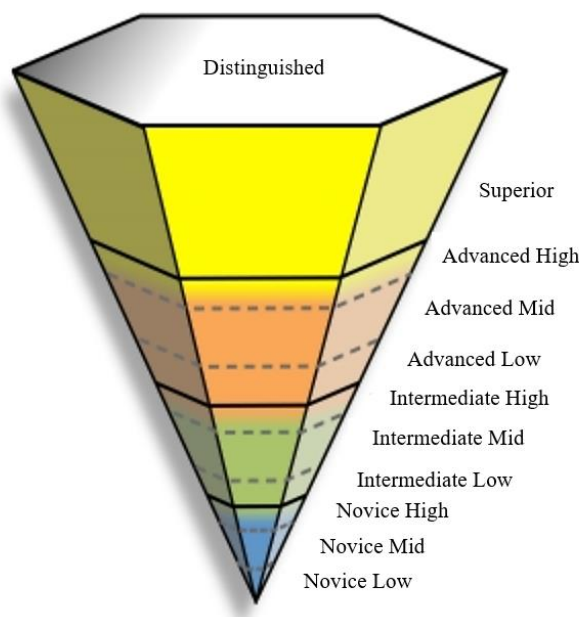


圖 1：ACTFL 大綱框架 (ACTFL Proficiency Guidelines 2012)

ACTFL 大綱中關於口語指標的內容, 首先是在 1986 年第一版中先提供了 N、I、A 三個主要等級, 以及 N、I 其下各三個次級, 共三等七級的內容; 再於 1999 年版本補充 A 等其下三個次級以及 S 等的內容, 使 ACTFL 大綱具備了四等十級的描述; 最終於現行的 2012 年增修版新增 D 等的口語能力描

述，完備了五等十一級的各等級口語指標內容。各等級口語指標內容皆以未經事先準備的真實語言情境為制訂原則，除了明示說話者在主要等級能完成的溝通任務，以及各任務相關的內容、語境、準確程度及談話類型之外，還簡略說明了說話者試圖達到下一個更高的主要等級時的能力侷限。這些口語指標能用來評估表達單向、非互動的演示型發言，或是互動、雙向交流的人際溝通（詳見 ACTFL 官網<sup>2</sup>）。

### 2.1.3 語言測驗與 CEFR、ACTFL 大綱框架之對應研究

近十多年，基於受測者了解施測結果意義、語言教師安排課程或建立分班標準、學習者規畫學習歷程，以及測驗研發者確保測驗信效度等各類實用需求，語言測驗成績與 ACTFL 大綱或 CEFR 的連結、不同語言框架的對應，成為越來越多研究者、測驗編制者關注的焦點。如 CoE 於 2009 年推出《Relating Language Examinations to the Common European Framework of Reference for languages: Learning, Teaching, Assessment-A Manual》（以下簡稱《手冊》）<sup>3</sup>，提出對應研究中可用以確保研究有效性之對應四步驟：熟悉 (familiarization)、說明 (specification)、標準設定 (standard setting) 以及效度驗證 (validation)。諸多投入對應研究的學者普遍遵循這四個步驟，如 Bärenfänger 和 Tschirner (2012)、Swender、Tschirner 和 Bärenfänger (2012)、American Council on the Teaching of Foreign Languages (2016)；但完整遵循四步驟進行對應研究需耗費較多資源及經費，因而 Jones、Ashton 和 Walker (2010) 認為《手冊》應鼓勵研究者可改革或彈性運用對應四步驟，有些學者則據此建議僅採用一至二個步驟進行對應，如鹿士義 (2011a, 2011b)、黃婷和賈國棟 (2012)、王昕 (2014)、王伯韜 (2015) 等；其中以標準設定步驟最常被擇用。

就主題來看，語言對應研究可概分兩類，第一類是將語言測驗對應至 CEFR 或 ACTFL 大綱等架構，以取得測驗通過門檻分數，或探究測驗分數所對應的語言能力等級，如 Tannenbaum 和 Wylie (2008)、Papageorgiou 等人 (2015)、鹿士義 (2011a, 2011b)、黃婷和賈國棟 (2012)、劉靜觀 (2012)、王昕 (2014)、王伯韜 (2015)、國家華語測驗推動工作委員會 (2015, 2016,

<sup>2</sup> ACTFL 大綱：<https://www.actfl.org/resources/actfl-proficiency-guidelines-2012>

<sup>3</sup> 《手冊》之下載網址：<https://www.coe.int/en/web/common-european-framework-reference-languages/relating-examinations-to-the-cefr>

2018)、羅蓮(2017)等。ACTFL 則於 2014 年進行多項 ACTFL 測驗對應至 CEFR 的研究，並依據研究結果於 2017 年更新《NCSSFL-ACTFL Can-Do Statements》<sup>4</sup>。

第二類研究屬於不同語言架構之間的對應關係，研究方法大致分為兩種，其一是分析並比較不同語言架構中所描述的語言功能與活動之異同，藉以建立彼此等級的連結關係；其二是透過語言測驗結果或學習者之口頭、書面語料為媒介，連結不同語言架構間的等級對應關係，本研究可歸為此類研究。前者如 American Council on the Teaching of Foreign Languages (2016) 分析並比較 CEFR 與 ACTFL 大綱中聽讀、口寫指標及相關量表的內容，取得彼此對應結果，並歸結出輸入型的聽讀能力有一致的對應結果，輸出型的口寫能力另有一致的對應結果，如 CEFR 聽讀最高等級 C2 對應至 ACTFL 大綱最高等級 D，但 CEFR 口語最高等級 C2 所對應的是 ACTFL 大綱的次高等級 S。同年，負責 ACTFL 各項測驗的 Language Testing International (2016) 也採用了對應能力指標的做法，透過對照 CLB 與 CEFR、ACTFL 大綱的任務及語言特徵等內容，分別就聽、讀、說、寫四項語言能力進行對應，從 CLB 4 至 10 級與 CEFR、ACTFL 大綱等級的對應結果，間接取得 CEFR 與 ACTFL 大綱各等級的對應關係。後者則如 Martínez Baztán (2008) 自編 9 道西語口語試題，評量 35 名西班牙格拉那達大學語言中心西語學習者的 ACTFL 大綱等級、CEFR 等級，最後提出 ACTFL 大綱與 CEFR 各等級的粗略對應結果為，A1 高於 NL，主要與 NM 對應，A1 高則可對應至 NH；A2 對應範圍含括 NH 高、IL，並觸及 IM 門檻；B1 主要與 IM 對應，B1 高則可對應至 IH；B2 低對應至 IH 高，B2 整體主要與 AL 對應；C1 對應範圍則自 AL 高、AM，至大部分 AH；C2 低與 AH 高對應，整體主要對應至 S (Martínez Baztán 2008:479-497)；不過該研究當時所採用的 ACTFL 大綱為 2001 年增修版，該版本僅提供 NL 至 S 共四等十級的口語描述，尚無現行架構中的最高等級 D 之口語指標，是以研究結果可能因當時的等級架構與現行的 2012 年版不同而有所落差。Bärenfänger 和 Tschirner (2012) 遵循《手冊》四步驟，使用以 CEFR 為架構依據的德國 OPIc (Oral Proficiency Interview-computer) 樣本混合 ACTFL 官網提供的 S 等級樣本，以德語口語能力為觀察對象，進行 ACTFL

---

<sup>4</sup> 《NCSSFL-ACTFL Can-Do Statements》：<https://www.actfl.org/publications/guidelines-and-manuals/ncssfl-actfl-can-do-statements>

大綱與 CEFR 對應研究，所得研究結果為 NH 與 A1 相當、IL 與 A2 相當、IM 對應至 B1.1、IH 對應至 B1.2、AL 對應至 B2.1、AM 對應至 B2.2、AH 與 C1 相當，以及 S 已對應至 C2。同年，Swender、Tschirner 和 Bärenfänger (2012) 另觀察了閱讀能力的對應關係，以 88 名德國萊比錫大學學生為研究對象，這些學生參加兩項分別參考 CEFR、ACTFL/ILR 架構設計的閱讀測驗，分析受測結果後得到 ACTFL 大綱 I 等與 CEFR 之 A 等相當，ACTFL 大綱 A 等與 CEFR B 等相當，ACTFL 大綱 S 等與 CEFR C1 等級對應的研究結論。趙家璧、藍珮君和陳柏熹 (2016) 透過兒童華語文閱讀測驗，採 Bookmark 法進行標準設定，從華語文閱讀能力的角度取得 CEFR Pre-A1 約與 ACTFL 大綱之 NL、NM 相當，A1 與 NH、IL 相當，A2 與 IM、IH 相當的對應關係。

綜上所述，雖已有前人探討並比較 CEFR 與 ACTFL 大綱的等級對應關係，然先前研究多針對英語、西語、德語之口、讀或聽說讀寫之綜合能力，以華語能力為對應研究媒介之研究僅占少數，如趙家璧、藍珮君和陳柏熹 (2016)，而針對華語口語能力進行的 CEFR 與 ACTFL 大綱之對應研究，目前幾乎未見相關文獻。因而，本研究期望以華語文口語測驗為媒介，從華語口語能力的角度，為 CEFR、ACTFL 大綱間口語能力之連結關係提供實證研究成果。

## 2.2 華語文口語測驗與 CEFR 等級對應

### 2.2.1 華語文口語測驗簡介

華語文口語測驗為華測會自 2008 年開始研發之標準化語言能力測驗，採電腦化測驗方式；測驗架構參照當時 CEFR 通行之三等六級架構，將華語口語能力亦分成三等六級；於 2012 年推出首次正式考試至今，試題內容取材皆參照 CEFR 對各等級外語學習者可處理情境範圍之描述，凡真實生活情境中，所有與溝通任務相關之主題，如個人資料及喜好、包含學校或職場情境在內之日常生活與休閒活動、旅行、購物、身體與健康、教育、居住環境含氣候與季節等皆涵蓋在內，不以特定教材為命題依據。2017 年初因應零起點學習者需求，向下新增一等級：準備級，同樣依測驗成績可再細分為準備級一級、準備級二級。如緒論所言，因已透過標準設定程序確認了準備級一級、準備級二級可分別對應至 ACTFL 大綱 NL、NM 兩次級 (國家華語測驗推動工作委員會 2017)，故本研究僅將入門基礎級、進階高階級與流利精通級列入研究範圍，以下也僅就此三等六級擇要概述。

測驗研發過程中，研發人員首先依據預計在各等級測得的口語能力，從 CEFR 提供的口語溝通任務項目中篩選出適用者，如表 1 所示。

表 1：華語文口語測驗研發依據之 CEFR 口語溝通任務項目列表

CEFR 指標項目	Band A	Band B	Band C
一般語言範圍 (Council of Europe 2018:131)	V	-	-
綜合口頭表達 (Council of Europe 2018:69)	V	V	V
對觀眾講話 (Council of Europe 2018:74)	V	V	V
持續性獨白：敘述經驗 (Council of Europe 2018:70)	V	V	-
持續性獨白：呈現案例 (Council of Europe 2018:72)	-	V	-
綜合口說互動 (Council of Europe 2018:83)	-	-	V
正式討論及會談 (Council of Europe 2018:87)	-	-	V

各等級選擇對應的能力項目不盡相同，乃因敘述與公告的能力在進階高階級就已經充份發揮，因此流利精通級指標的制定，選擇以綜合口說互動、正式討論及會談等項目，取代入門基礎級、進階高階級測驗中持續性獨白（敘述經驗、呈現案例）與公告這三個項目；入門基礎級則因尚未具備能完整呈現公告與持續性獨白（呈現案例）的能力，因而納入一般語言範圍的能力描述做為能力指標的來源之一。整合上述從 CEFR 中擇用之溝通任務項目的描述後，研發人員即據此擬定華語文口語能力的各等級指標（見表 2）、測驗題型（見表 3）。

入門基礎級階段的語言使用者，發展的口語表達能力為描述性能力，由入門級時「能使用熟悉的日常用語與詞彙，進行簡短地敘述」，發展至基礎級時，能進一步「使用簡單短語或句子，運用簡單連接詞或條列方式，進行連貫描述」；據此大致可將入門基礎級階段語言學習者的口語表達能力分為「單句層次描述能力」和「段落層次描述能力」二大面向，該等級題型規劃為單句簡答類的回答問題 3 題和描述類試題 3 題，描述類試題再依試題素材細分為經驗描述題 3 題、影片描述題 1 題。進階高階級，使用者除描述能力大幅度發展，論證說明能力也逐漸萌芽發展，因而此階段的口語能力可大致分為「描述能力」和「說明能力」二大面向，並分別規劃為描述類和說明類兩大類題型：描述類試題依試題素材細分為經驗描述題 2 題、圖片描述題 1 題；說明類題型則細分為選擇並比較不同方案、提出建議，以及對特定議題表達



贊成或反對立場等三種能力的陳述意見題。流利精通級使用者的口語表達能力主要展現在具有整合各種資料來源及運用回應、論辯等方式充分且堅定地表達個人觀點，因此，主要區分為「摘要能力」、「回應能力」以及「論述能力」等三大面向，據此，則分別規劃了回應類、摘要類及論述類等三大題型；其中，回應類題型採取類似多輪對話的辯論模式，設計為一組由角色扮演題一題及觀點回應題兩題所組成的題組，除用以考察受測者之回應能力，也盼能達到類似 OPI (Oral Proficiency Interview) 測驗模式的互動溝通功能；至於摘要暨觀點論述類試題，則著重於評量受測者能否「依據題目提供的文章，歸納主題及其相關子題的多方重點，並整合出一段持續性且融合多方觀點的統整性口頭摘要」。

表 2：華語文口語測驗能力指標

TOCFL 等級	CEFR 等級	華語文口語能力指標
精通級 Level 6	C2	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. 能游刃有餘地以清楚、具說服力的論證維持立場。 3. 彈性地調整說話方式以符合聽者需求。
流利級 Level 5	C1	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。
高階級 Level 4	B2	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。
進階級 Level 3	B1	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。

表 2：華語文口語測驗能力指標（續）

TOCFL 等級	CEFR 等級	華語文口語能力指標
基礎級 Level 2	A2	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。
入門級 Level 1	A1	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。

表 3：華語文口語測驗題型簡介

測驗等級	題型	題型說明及測驗目標	題數
入門 基礎級 Band A	回答 問題	【測驗目標：回答能力】 需以完整、簡短的句子回答問題。	單題 (4 題)
	經驗 描述	【測驗目標：描述能力】 需依據設定的情境，描述所處環境中熟悉的日常事物或學習經驗。	單題 (3 題)
	影片 描述	【測驗目標：描述能力】 提供一段約60秒的影片，影片會播放兩次，需以完整、簡短的句子描述影片內容。	單題 (1 題)
進階 高階級 Band B	經驗 描述	【測驗目標：描述能力】 需依據題目設定的情境，描述個人的經驗、事件、夢想等。	單題 (2 題)
	圖片 描述	【測驗目標：描述能力】 提供一張圖片，需描述圖片內容，並回答一個相關的問題。	單題 (1 題)
	陳述 意見	【測驗目標：說明能力】 需依據設定的情境提出看法，使用相關說明及適當例子，解說並支持個人論點。	單題 (3 題)

表 3：華語文口語測驗題型簡介（續）

測驗等級	題型	題型說明及測驗目標	題數
流利 精通級 Band C	角色 扮演	<b>【測驗目標：回應能力】</b> 需依據題目設定的情境及角色，使用適切的說話方式，針對議題發表與所需扮演之角色立場一致的看法並做出結論。	題組
	觀點 回應	<b>【測驗目標：回應能力】</b> 需依據題目設定的情境和角色，回應相反論證的看法，並維持指定角色的立場。	
	文章 摘要	<b>【測驗目標：摘要能力】</b> 提供兩篇主旨相關的文章，每篇約500字；需依據兩篇文章的內容歸納重點，並用自己的話重新組織內容，進行口頭摘要。	題組
	觀點 論述	<b>【測驗目標：論述能力】</b> 需依據設定的情境發表觀點，並提出論證以維持個人立場。	
	陳述 意見	<b>【測驗目標：論述能力】</b> 需依據設定的情境提出看法，並使用相關說明及適當例子來解說、支持個人論點。	單題 (1 題)

### 2.2.2 透過標準設定程序制定華語文口語測驗通過門檻

華語文口語測驗各等級之通過門檻係採用標準設定程序所設定。由於入門基礎級、進階高階級與流利精通級的測驗給分方式採多元計分制 (polytomous items)，與單選題非對即錯的概念不同，是以通過門檻設定方法參考 Yes/No Angoff 法之概念，再因應測驗形式為建構反應題加以調整，標準設定程序各步驟及細節可參考藍珮君等人 (2013)。於 2013 至 2016 年間先後完成進階高階級、入門基礎級以及流利精通級對應至 CEFR 之研究，標準設定成員均由華語文及語言學領域專家所組成，並依循標準化流程執行，最後依據研究成果確立了各等級與 CEFR 對應關係，即入門級、基礎級、進階級、高階級、流利級以及精通級，依序可分別對應至 CEFR 的 A1、A2、B1、

B2、C1 以及 C2（國家華語測驗推動工作委員會 2015, 2016, 2018）。

完成測驗通過門檻設定後，需檢視標準設定結果是否可靠且具效度。標準設定結果的效度檢核一般可分為程序性效度、內部效度及外部效度三部分 (Kane 1994)，此處提供的是程序性效度及內部效度檢核結果。首先，程序性效度方面，每次標準設定會議皆按照既定議程進行，且在各回合間給予與會者充分的討論時間。會議後的問卷調查結果顯示，與會者均同意會議帶領者對會議目的及任務解釋清楚、對標準設定方法的操作流程說明得很清楚、能了解最低能力者在標準設定方法的涵義、每回合後團體討論和分享，有助於進行下一回合的判斷、對於自己所設定的切截分數有信心等等，可做為程序性效度依據。其次，內部效度證據部分則可由兩方面來看，分別是每一回合通過門檻的標準差，以及每一回合音檔 CEFR 等級判斷與實際級分之斯皮爾曼等級相關。標準差部分，入門基礎級、進階高階級與流利精通級三個測驗等級六個通過門檻的標準差均為第一回合最高，然後逐漸降低，顯示經過幾個回合的討論後，與會者的判斷漸趨一致。另外，入門基礎級與進階高階級第三回合各題型間的相關係數分別介於 .881 至 .937 之間 ( $p < .01$ ) 與 .703 至 .949 之間 ( $p < .01$ )；顯示專家們對於音檔通過等級的判斷，與實際得分之間具有中高度或高度的正相關，判斷結果與實際得分頗為一致。結論是，華語文口語測驗標準設定結果在程序性效度與內部效度二項證據均獲得支持。

### 2.3 標準設定方法與選擇

標準設定，指的是建立測驗通過分數的過程 (Cizek 2006)，也就是適當地依循規定好的、合理的規則或程序，達成指定一個數字以區辨兩種或兩種以上的表現程度 (Cizek 1993)。換言之，標準設定是一種標記，代表常用於判斷測驗分數的方式，而這個測驗分數可以支持測驗對受測者知識、技巧、能力、精熟度或準備度等水準的決定 (Tannenbaum and Wylie 2008)。

目前大型標準化測驗，如美國教育測驗服務社 (Educational Testing Service, ETS) 研發的托福測驗 (TOEFL)，普遍採用標準設定程序進行測驗與語言能力指標的對應 (Tannenbaum and Wylie 2008)。CEFR 也在《手冊》中，明確指出標準設定步驟可有效確保對應研究的有效性，將其列為對應研究建議採取的四步驟之一。因此本研究希望邀集華語教學、測驗及語言學相關領域的專家，透過標準設定程序以了解採用 CEFR 架構之華語文口語測驗與

ACTFL 大綱之間的對應關係。

梳理文獻後，發現標準設定的研究方法十分多元，適用於多元計分題型的標準設定方法有美國國家教育進展評量 (National Assessment of Educational Progress, NAEP) 於 1992 年使用之 Exemplar Response 法 (Jaeger et al. 1996) 與 Lewis、Mitzel 和 Green (1996) 提出的書籤法 (Bookmark Method)。這兩種設定方法的共同處在於，皆是經由專家小組會議進行難度判定與討論的程序化過程，做為測驗訂定通過標準的方法；差異處有二：其一是判斷方式的不同，實施於口語測驗，Exemplar Response 法要求參與者從依據不同級分挑選之受測者音檔中，挑選出符合描述的音檔；書籤法則是參與者需在一系列依表現程度由低至高排列的受測者樣本音檔中，標記出最低能力者具有特定答對率的單一音檔。其二在於是否採用試題反應理論 (Item Response Theory, IRT) 參數，書籤法在實施上需將不同受測者樣本音檔依其程度高低置於同一量尺上排序，因此常採用 IRT 參數，Exemplar Response 法則因為是逐一判斷各音檔，而不一定需要使用 IRT 參數 (趙家璧、藍珮君、陳柏熹 2016)。

由於此次研究之與會者是透過華語文口語測驗各級分音檔進行 ACTFL 大綱各等級語言能力描述之對應研究，而華語文口語測驗之三個測驗等級雖已採水平等化，但未進行跨等級間的垂直等化，無法將不同等級不同級分的音檔置於同一量尺上比較，若採用標籤法，施行上較為困難，是以在評估後，本研究決定採用 Exemplar Response 法做為主要研究方法。

### 3. 研究方法

本研究採用 Exemplar Response 法進行對應，過程包括標準設定會議的前置作業 (如挑選本研究適用的標準音檔)、執行標準設定會議，以及會後分析標準設定結果等三個主要階段。本節將先簡述此次研究使用的標準音檔、與會者背景以及等級標記方式，再介紹本研究執行標準設定程序的詳細步驟。

#### 3.1 研究工具

本研究採用華語文口語測驗入門基礎級、進階高階級、流利精通級各題型各級分之受測者標準音檔為研究工具：入門基礎級 0 至 3 級分標準音檔，共 8 個；進階高階級 1 至 5 級分標準音檔，共 10 個；流利精通級 1 至 5 級分標準音檔，共 15 個；總計 33 個音檔。音檔編號方式為「測驗等級-測驗級分

-該等級題型的代號-音檔序號」，如「A-0-2-1」即代表入門基礎級 0 級分描述題的第一個音檔，「B-5-2-1」代表的是進階高階級 5 級分陳述意見題的第一個音檔，同理，「C-4-3-1」代表流利精通級 4 級分觀點論述的第一個音檔。33 個音檔在各等級內部為由低而高的分布，音檔排列的原則可參見圖 2。這些標準音檔為經過訓練的評分者共同挑選之典型音檔，足以代表各級分的典型口語能力表現。會議中還提供了音檔之試題內容，輔助與會者判斷音檔內容與 ACTFL 大綱中口語指標的符合程度。

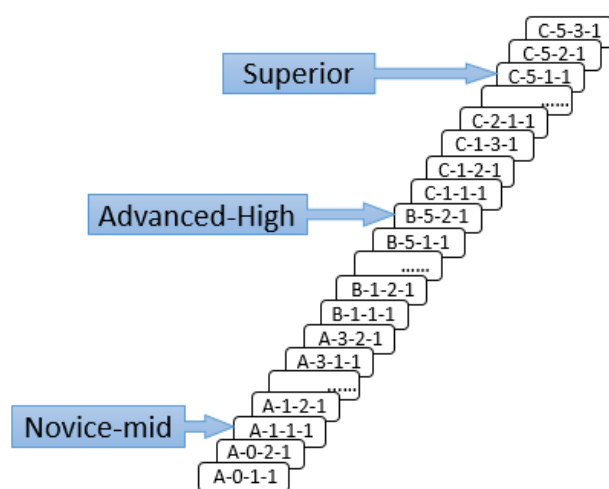


圖2：口語音檔排列示意圖

### 3.2 操作方法

本研究採用 Exemplar Response 法判定 ACTFL 大綱口語指標自初級至優異級，各等級之受測者級分。原則上，無論採用何種方法，在開始進行標準設定研究前，通常需先討論並制訂出測驗通過門檻之最低能力描述，若測驗可分為數個等級，就需制訂數個最低能力描述。然而，ACTFL 大綱的五個主要等級共涵蓋十一個次級，分級多且精細，實難以從某等級的描述中再定義「最低能力者」表現，為確保與會者能對判定標準掌握高度一致性，研究者和測驗專家經討論後，將研究方式調整為把各等級口語描述視為典型能力描述，而與會者所判定的音檔編號，即代表最接近該等級典型能力者的口語表現。

實際操作方式為提供入門基礎級、進階高階級與流利精通級這三個等級的各級分標準音檔，請與會者依照 ACTFL 大綱的口語指標，判斷符合描述

的口語表現最接近哪個標準音檔。因華測會已另完成準備級口語測驗之標準設定研究（國家華語測驗推動工作委員會 2017），確認 NL、NM 分別對應至準備級一級與準備級二級，而因 NH 須以兩相鄰等級中間值求最低門檻（見 3.5 節），仍需對 NM 進行判定。所以，此次對應的 ACTFL 大綱範圍為 NM、NH、IL、IM、IH、AL、AM、AH、S 及 D，共十個等級。

判斷過程中，最低的 NM 與最高的 D 等級，若與會者認為沒有任何一個音檔符合該等級描述，則可選擇不標記。此外，因華語文口語測驗未對跨測驗等級進行垂直等化，故每個等級的最高級分不易與下一個更高等級的最低級分直接比較，若與會者判斷某一等級時，認為不同測驗等級分別有音檔同時符合 ACTFL 大綱口語指標，則兩個音檔編號皆可標記於同一等級的細格內，如：將 A-3-2-1 和 B-2-1-1 兩個音檔編號都標記在 IM 細格內。每一回合判斷完畢後，研究者會將所有與會者的判斷結果彙整如圖 3。

音檔級分 ACTFL 等級	A_0	A_1	A_2	A_3	B_1	.....	B_5	C_1	.....	C_5	Min.	Max.	Median
Novice-Mid	8	7	1	0	0	...	0	0	...	0	A_0	A_2	A_0
Novice-High	0	13	2	1	1	...	0	0	...	0	A_1	A_3/ B_1	A_1
Intermediate-Low	Novice-High最低能力對應到 華語文口語測驗入門基礎級 (Band A) 的0級分至1級分之間										A_2	B_1	A_2

圖 3：ACTFL 大綱口語指標與音檔判斷結果範例

### 3.3 研究與會者

本次標準設定會議於 2018 年 9 月 15 至 16 日舉行，與會者專業背景概分為資深華語教師以及語言教學領域學者這兩類，共 17 人（表 4）：具資深華語教師身分者 13 人，除了皆擁有任教於國內華語教學中心或多年海外華語教學的經驗，教學年資介於 9 至 35 年，還分別具備華語師資培訓講師、初級華語或商業華語教材編撰等經歷，且其中 5 位具 ACTFL OPI 經驗或曾受過 OPI 評分訓練，熟悉 ACTFL 大綱能力指標；另有 8 位與會者具有華語文口語測驗評分經驗達 5 年以上。語言教學領域之學者為 4 位，分別來自語言學、漢語語法、華語教學、數位教學等學術領域。所有與會專家中，女性有 14 位，男性有 3 位。標準設定會議時間為兩天，由於其中 2 名與會者各僅能參與一天，故兩天會議實際與會的人數各為 16 人。

表 4：與會者背景分布表

與會者背景		人數	百分比
(1) 專業背景	資深華語教師	13	76.5%
	語言教學領域專家	4	23.5%
(2) OPI 相關經驗	有	5	29.4%
	無	12	70.6%
(3) 性別	女	14	83.4%
	男	3	17.6%

### 3.4 研究步驟

本研究執行之標準設定程序步驟如下：

- (1) 會議召開前兩週先將 ACTFL 大綱的口語指標內容、本研究動機及研究方法介紹等資料提供給與會者，以利與會者先行熟悉能力指標與標準設定流程。
- (2) 會議為期兩天，第一天進行 NM 至 AL，第二天進行 AM 至 D。皆先簡介標準設定的目的及流程、Exemplar Response 法操作方式。
- (3) 說明 ACTFL 大綱的分級方式與口語指標，說明中亦強調相鄰等級或次級間的口語能力差異。
- (4) 介紹華語文口語測驗的等級架構及題型，含題型設計方向與受測者應試時的作答方式。
- (5) 與會者進行第一回合標記，從 33 個華語文口語測驗標準音檔中，根據 ACTFL 大綱各等級口語指標，於聆聽音檔後，挑出與各等級指標最符合的音檔，標記音檔編號並寫下判定原因。
- (6) 將步驟 (5) 判斷結果所得之回饋訊息提供給與會者，包括：(a) 每位與會者第一回合的標記結果；(b) 透過判斷結果和音檔給分資料所得出的各等級級分描述統計數值，包括最小值、最大值、眾數與中位數；(c) 判定等級的暫定通過門檻。與會者根據上述回饋資訊，依等級順序由 NM 開始討論。
- (7) 完成第一回合討論後，與會者進行第二回合標記，判斷方式同步驟 (5)。
- (8) 發放問卷調查與會者對當天標準設定會議內容與程序的了解和判定過程的掌握度。



### 3.5 資料分析

首先，研究者輸入 17 名研究與會者於第一回合及第二回合 ACTFL 大綱之 NM、NH、IL、IM、IH、AL、AM、AH、S 及 D 這十個等級口語指標與華語文口語測驗標準音檔的判斷結果，亦即各等級典型口語表現的音檔等級與級分，接著計算出第一回合與第二回合與會者判定結果的最小值、最大值與中位數。標記方式為音檔所代表的華語文口語測驗等級與級分，如「A\_2」表示「入門基礎級 (Band A) 的 2 級分」，以此類推。中位數是指將所有數字依大小順序由小至大排列後，排列在最中間的數字。相較於平均數，中位數對極端值不敏感，受極端值的影響較小。

接著，再根據十個等級第二回合的典型口語指標表現判定音檔之中位數，計算各等級單題最低門檻，計算方式為相鄰兩等級典型樣態級分的中間值。如：NM 與 NH 典型樣態級分的平均數即為 NH 的單題最低門檻。單題最低門檻再乘以與其對應之華語文口語測驗之測驗等級總題數，所得之級分總和始為該 ACTFL 大綱等級通過門檻；其餘等級以此類推。

在與會者判斷結果的一致性方面，本研究提出的內部效度證據有二：一是與會者於各等級標記音檔級分之分布情形；二是問卷調查結果的分析，計算與會者對問卷各題的平均同意程度，勾選「非常同意」者標記為 4，「同意」為 3，「不同意」為 2，「非常不同意」為 1；再換算為同意百分比，藉此了解與會者對於整個標準設定流程的理解程度。

## 4. 研究結果

由於 15 號與會者在第一天會議第一回合與第二回合的標記結果與其他與會者相差較大，六個等級中有三個等級標記音檔為所有與會者中最高，且有一個等級標記音檔級分較前一等級低，顯示其不甚理解操作方式，研究者判斷此筆資料會對結果分析造成影響，故刪除其第一天標記資料，以下結果皆從其餘 16 位研究與會者的標記資料分析而來。

### 4.1 標準設定結果

表 5 為第一回合與第二回合與會者判定各等級典型樣態級分的描述統計結果，由表 5 可知，經過第二回合的判定後，在級分中位數的部分，除了 IL 的級分較第一回合降低，由入門基礎級 3 級分降至 2 級分，以及 D 由第一回合之流利精通級 5 級分，調整為該音檔之口語表現不足以代表 D 的典型口語

能力，而未能取得可對應的 CEFR 等級之外，其餘八個等級的級分中位數都與第一回合相同。整體而言，與會者在第一回合與第二回合對各等級的判斷變化不大，代表本研究與會者對於各等的語言能力有大致相同的理解與判定標準。進一步計算 ACTFL 大綱各等級單題最低門檻結果如下：

NH 最低門檻介於入門基礎級 0 級分與 1 級分之間，IL 則介於入門基礎級 1 級分與 2 級分之間，IM 門檻落在入門基礎級 2 級分與 3 級分之間，IH 門檻在進階高階級 2 級分與 3 級分之間，AL 介於進階高階級 3 級分與 4 級分之間，AM 介於進階高階級 4 級分與 5 級分之間，AH 門檻位於流利精通級 2 級分與 3 級分之間，S 最低門檻落在流利精通級 3 級分與 4 級分之間，D 則因與會專家判斷流利精通級 5 級分音檔尚不足以代表此等級典型口語表現，而無法求出對應之測驗等級與級分。

至於整卷最低門檻，計算方式為單題最低門檻級分乘以題數。以 NH 為例，NH 門檻在入門基礎級 0 級分與 1 級分之間，因華語文口語測驗入門基礎級共有 8 題，故 NH 的整卷最低門檻為入門基礎級測驗 4 分(= $(0+1)/2$ )\*8)，其餘等級之整卷最低門檻級分如下表所示。

表 5：第一回合與第二回合與會者判定各等級典型樣態級分描述統計結果

ACTFL 大綱等級	第一回合			第二回合			整卷 門檻	CEFR 通過門檻
	最小值	最大值	中位數	最小值	最大值	中位數		
NM	N/A	A_2	A_0	A_0	A_0	A_0	N/A	--
NH	A_1	A_3/B_2	A_1	A_1	A_2	A_1	A_4	A_8(A1) A_16(A2)
IL	A_2	B_3	A_3	A_2	B_1	A_2	A_12	
IM	A_2	B_4	B_2	A_3	B_2	A_3/B_2	A_20	B_12(B1) B_24(B2)
IH	A_2	B_5/C_2	B_3	B_2	B_4	B_3	B_15	
AL	B_2	C_3	B_4	B_3	B_5/C_2	B_4	B_21	C_21(C1) C_35(C2)
AM	B_4	C_3	B_5/C_2	B_4	C_3	B_5/C_2	B_27	
AH	B_5/C_2	C_4	C_3	B_5/C_3	C_4	C_3	C_17.5	C_21(C1)
S	C_3	C_5	C_4	C_4	C_5	C_4	C_24.5	C_35(C2)
D	C_5	N/A	C_5	C_5	N/A	N/A	N/A	--

#### 4.2 標準設定結果之效度

完成標準設定程序後，需檢視標準設定結果是否可靠、具有效度。一般來說，標準設定結果的效度檢核可分為程序性效度、內部效度，以及外部效

度三部分 (Kane 1994)，此處提供程序性效度及內部效度檢核結果。

程序性效度方面，華語文口語測驗對應至 ACTFL 大綱口語指標的標準設定過程，均依照既定步驟，與會者在判定後的討論亦進行充分討論與分享，如第一回合標記後的討論約進行兩至三小時，交流彼此的判定依據。且兩天會議的問卷調查結果（見表 6）各題平均數都達到 3.4 以上，同意百分比達 85% 以上；18 題中，有 6 題平均數達 4.0，同意百分比高達 100%。此結果顯示與會者了解標準設定流程與 Exemplar Response 法進行方式，也了解 ACTFL 大綱各等級口語指標的意涵等研究相關資訊，以上皆顯示標準設定結果具有程序性效度。

表 6：標準設定會議問卷調查結果<sup>5</sup>

問卷題目	平均數		同意者百分比	
	D1	D2	D1	D2
1. 我了解本次 TOCFL 與 ACTFL 大綱口語指標對應研究會議的目的。	4.0	4.0	100%	100%
2. 會議帶領者對於標準設定研究的方法、流程說明得很清楚。	4.0	4.0	100%	100%
3. 會議帶領者對於本研究 Exemplar Response 法的進行方式說明得很清楚。	3.9	4.0	98%	100%
4. 會議帶領者對於 ACTFL 大綱的架構和口語指標說明得很清楚。	3.8	3.9	95%	98%
5. 我了解 ACTFL 大綱各等級口語指標的意涵。	3.7	3.9	92%	97%
6. 研究者提供各等級依能力程度排序的音檔，符合我認為的能力相對程度。	3.4	3.6	85%	89%
7. 我是根據 ACTFL 大綱各等級口語指標，來標記符合口語指標表現的音檔。	3.7	3.9	92%	98%
8. 第一回合後團體中的討論和分享，有助於我進行第二回合的判斷。	3.9	4.0	97%	100%
9. 我對於自己所設定的切點結果 (cut score) 有信心。	3.6	3.8	91%	94%

<sup>5</sup> 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。D1 表示第一天，D2 表示第二天。第 4 題至第 9 題問卷題目在 D1 和 D2 所指涉的等級不同，D1 是指 N、I、A；D2 是指 A、D、S。

內部效度則由兩個回合與會者各等級所標記音檔級分之分布情形作為依據，由於音檔涵蓋級分分布廣泛，以下分兩表呈現，表 7 為 NM 至 IH，表 8 為 AL 至 D。表 7 可見，NM、NH、IL 與 IH 的級分分布在第一回合均較為分散，經過第一回合討論後，第二回合的級分分布皆較第一回合集中，且八成以上的與會者判斷相同級分。以 NH 為例，第一回合分散在六類級分(A\_1、A\_2、A\_3、A\_3/B\_1、A\_3/B\_2、B\_1)，第二回合則有 93%的與會者判斷為 A\_1，其餘 7%與會者判斷為 A\_2。而 IM 兩個回合同為 47%與會者判斷相同級分(B\_2)，但級分的分布情形從六類縮小至三類。表 8 亦是類似情形，AL、AH、S 與 D 在第二回合的級分分布情形均較第一回合集中，且有 50%以上的與會者判斷相同級分。AM 部分，兩個回合與會者的判斷結果都分散在相同的五類，較高比例者均為 B\_5/C\_2，但兩回合比例皆未過半。

表 7：初級中等 (NM) 至中級高等 (IH) 與會者判斷音檔級分分布表

		A_0	A_1	A_2	A_3	A_3/B_1	A3/B_2	B_1	B_2	B_3	B_4	B_5	B_5/C_1	B_5/C_2	N/A
NM	R <sub>1</sub> <sup>6</sup>	47%	20%	27%											7%
	R <sub>2</sub>	100%													
NH	R <sub>1</sub>		53%	7%	7%	13%	13%	7%							
	R <sub>2</sub>		93%	7%											
IL	R <sub>1</sub>			47%	7%	7%		13%	20%	7%					
	R <sub>2</sub>			80%	7%			13%							
IM	R <sub>1</sub>			13%	7%		13%		47%	13%	7%				
	R <sub>2</sub>				27%		27%		47%						
IH	R <sub>1</sub>			7%			13%		7%	33%	27%	7%		7%	
	R <sub>2</sub>								7%	87%	7%				

<sup>6</sup> R<sub>1</sub> 表示第一回合，R<sub>2</sub> 表示第二回合。

表 8：高級初等 (AL) 至優異級 (D) 與會者判斷音檔級分分布表

		B_2	B_3	B_4	B_5	B_5/C_1	B_5/C_2	B_5/C_3	C_1	C_2	C_3	C_4	C_5	N/A
AL	R <sub>1</sub>	7%	13%	<b>40%</b>	7%	13%			7%		13%			
	R <sub>2</sub>		7%	<b>87%</b>			7%							
AM	R <sub>1</sub>			6%	19%		<b>44%</b>			19%	13%			
	R <sub>2</sub>			6%	31%		<b>38%</b>			13%	13%			
AH	R <sub>1</sub>						6%	6%			<b>50%</b>	38%		
	R <sub>2</sub>						6%				<b>56%</b>	38%		
S	R <sub>1</sub>										6%	<b>81%</b>	13%	
	R <sub>2</sub>											<b>94%</b>	6%	
D	R <sub>1</sub>												<b>63%</b>	38%
	R <sub>2</sub>												38%	<b>63%</b>

上述結果顯示，經過討論後與會者間的一致性提高，十個等級中有九個等級級分分布縮小，八個等級有超過半數的與會者判斷相同級分，可做為標準設定結果的內部效度證據。進一步檢視第二回合與會者判斷相同級分比例未達 50% 的 IM 與 AM，研究者推測可能因為這兩個等級剛好位於華語文口語測驗三個測驗等級的交界：以 IM 為例，前一個等級為 IL，80% 與會者判斷音檔級分為入門基礎級 2 級分 (A\_2)，後一個等級為 IH，87% 與會者判斷音檔級分為進階高階級 3 級分 (B\_3)。第一回合結果，與會者判斷 IM 的音檔級分分布涵蓋了入門基礎級與進階高階級，共有六類；第二回合縮小至三類，分布在 A\_3、A\_3/B\_2、B\_2。至於 AM 判斷級分則涵蓋進階高階級與流利精通級。跨不同測驗等級的音檔在試題的主題範圍與需完成任務均不同，可能造成與會者判斷不易，而影響整體判斷的一致性。

#### 4.3 各等級之判別與討論焦點

與會者在音檔等級判定的討論中，都需表述個人的判定歷程與策略。

在採用 ACTFL 大綱各等級口語指標中的典型特徵，對比試題文本內容及口語音檔中的內容組織、語言運用、話語表現等向度的能力表現時，幾乎所有與會者都表示，若他們注意到具某等級典型的特徵開始出現在某口語音檔中，便會進一步觀察該特徵是否在相同或相鄰級分的音檔間密集地成為答題的典型能力，若某級分區段的音檔表現都具備某個或某些典型能力特徵時，

該級分區段的前後級分相關音檔很可能成為研究與會者定下等級標記的切點，或成為稍後討論中較容易被討論的位置。

以NH等級的討論內容為例，與會者們提出：

「A-1-1-1、A-1-2-3這兩個音檔都使用短句或短語，詞彙、語法基本，錯誤不少，準確度低，不易理解。很明顯是高度依賴學過的短語，重複也多，這是可判斷為NH等級的重要依據。」～與會者10

「A-1-2-3這個音檔，符合NH等級回應簡單直接的問題、可預測活動，也都是用學過的短語、短句簡單回答，重複多、也常錯誤。」～與會者01、與會者06

總結與會者將音檔判定為 NH 等級的典型特徵，在於「話題可預測」、  
「多使用短句或高度依賴學過的短語、常重複且準確性較低」等，凡呈現上述口語表現的音檔內容，就會被判定為 NH 等級。

再以 AL 等級的討論內容為例，與會者們提出：

「我認為B-4-4-1音檔最接近ACTFL大綱AL等級，判斷依據在於該音檔符合AL等級的敘述，問題的主題本來就是Advanced等級的任務情境，這個音檔也做到了結構完整，能組織成段落，也具備清楚的時間框架；再來，詞彙較豐富多樣，但有的準確性不高，這些都是很符合AL等級的描述.....」～與會者10

「B-4-1-1音檔的敘述內容結構完整、成段、語序事件時間的句子完成度高，詞彙用的很豐富，雖然錯誤多，穩定性不高，但基本的也都用對，但這就是AL這個等級的能力描述談到的關鍵.....」～與會者06、07

總結與會者將音檔判定為 AL 等級時的關鍵，在於即便主題涉及稍具複雜度的具體話題，話語內容也能結構完整、發展出清楚的時間架構，以及能將長句組合為段落話語，能掌握一般句法的基本結構，詞彙豐富但準確度不高。

前述 NH 及 AL 與會者提出之判斷特徵相似，同時此兩等級分別在表 7、表 8 之第二輪判斷的音檔級分分布一致性較高，達到 93%與 87%，顯見等級討論焦點亦可作為標準設定結果之效度證據。

#### 4.4 指標對應結果與等級難度對應

經過兩天對應研究會議的討論與設定，本研究以華語文口語測驗為媒介，取得 ACTFL 大綱 NM 至 S 九個等級最低能力者於三個測驗等級的級分對應關係，即表 5 之整卷門檻級分，再與口語測驗對應 CEFR 架構 A1 至 C2 等級通過門檻分數進行彙整。先前提及因口語測驗跨測驗等級間未進行垂直等化，是故對應圖分為入門基礎級、進階高階級與流利精通級三個區塊，各自依照測驗等級的級分範圍呈現 CEFR 與 ACTFL 的對應關係。結果如圖 4 所示，CEFR A1 程度的華語口語能力介於 ACTFL 大綱的 NH 至 IL；A2 程度位於 IL 與 IM 之間；B1 程度略低於 IH；B2 程度落在 AL 與 AM 之間；C1 程度介於 AH 至 S；C2 程度則高於 S；D 可能囿於音檔涵蓋能力範圍的限制，未能訂出對應位置。

TOCFL 口語	CEFR	ACTFL
流利精通級	C2	Superior
	C1	Advanced High
進階高階級	B2	Advanced Mid
		Advanced Low
	B1	Intermediate High
入門基礎級	A2	Intermediate Mid
		Intermediate Low
	A1	Novice High

圖 4：CEFR、華語文口語測驗與 ACTFL 大綱等級對照表（擷自華測會官網<sup>7</sup>）

<sup>7</sup> <https://tocfl.edu.tw/index.php/test/speaking/list/8>

## 5. 討論與建議

### 5.1 各等級典型樣態級分結果

如前節內容所述，本研究得出 CEFR 與 ACTFL 大綱對應關係為：CEFR A1 程度的華語口語能力，介於 ACTFL 大綱的 NH 至 IL 之間；A2 程度位於 IL 與 IM 之間；B1 程度略低於 IH；B2 程度落在 AL 與 AM 之間；C1 程度主要落在 AH 至 S 之間；C2 程度則高於 S；ACTFL 大綱最高等級 D 則未能訂出對應位置。

此結果大致與研究者分析兩大架構能力指標後的預設相當接近。而過往 CEFR 與 ACTFL 大綱對應的相關研究中，Martínez Baztán (2008) 為較早期的研究，當時 ACTFL 大綱中的口語指標架構僅四等十級，無現行架構之最高等級 D，該研究對應結果為 A1 高於 NL，主要與 NM 對應，A1 高則可對應至 NH；A2 對應範圍含括 NH 高、IL，並觸及 IM 門檻；B1 主要與 IM 對應，B1 高則可對應至 IH；B2 低對應至 IH 高，B2 整體主要與 AL 對應；C1 對應範圍則自 AL 高、AM，至大部分 AH；C2 低與 AH 高對應，整體主要對應至 S。其後，Bärenfänger 和 Tschirner (2012) 與 American Council on the Teaching of Foreign Languages (2016) 對應口語指標的研究結果皆為 NH 與 A1 相當、IL 與 A2 相當、IM 對應至 B1.1、IH 對應至 B1.2、AL 對應至 B2.1、AM 對應至 B2.2、AH 與 C1 相當，S 則對應至 C2。

整體而言，各研究在 CEFR 的 A1 至 B2 等級的對應結果大致相同，但在 C1、C2 等級的對應情況則落差較大。簡言之，CEFR 架構之 A1 至 B1 程度約相當於 ACTFL 大綱的初級至中級範圍，B2 約與 ACTFL 大綱之高級相當，C1 則分別對應到 ACTFL 大綱之高級高等或優秀級，C2 等級則對應至 ACTFL 大綱口語指標的優秀級或優異級。研究者推論可能原因如下：一是各研究所採用的對應材料類型不一致，有的採用文字型式的能力指標內容，有的採用實際口語音檔，因而造成對能力範圍認知與理解上的差異；其二可能因各研究用來對應的語言不同，無法完全推論到其他語言 (Council of Europe 2009; American Council on the Teaching of Foreign Languages 2016)；有待未來更多相關研究確認。

### 5.2 應用參考

本研究結果為 CEFR 與 ACTFL 大綱兩大架構的各等級口語指標做出客觀連結；同時，也在進行各等級設定的過程中，確認彼此在各等級口語能力中



具有的數項共通特徵，可供外語口語教學者、學習者及測驗研發者參考。

華語文口語測驗入門基礎級直接對應的 CEFR A1 至 A2 等級口語指標中提及，此階段使用者所能掌握的詞語、主題與語言使用情境都以日常例行生活等經驗且與個人熟悉度高相關者為範疇，且具有可預測性；上述描述與 ACTFL 大綱之 NH、IL 及部分 IM 的指標相似。口語表達能力發展歷程部分，也存在著共通性：皆由「能使用數個記憶限定的日常用語與詞彙簡短敘述」(A1↔NH)的單句層次描述能力進展至稍具穩定度的段落層次描述能力，也就是開始「能將數個簡單短語或固定句式組合成串，針對簡單、特定的可預測情境進行描述」(A2↔IL~IM)，且此階段的說話者具備頻繁停頓(A1~A2↔NH)、重組詞語(A1~A2↔NH~IL)等語言特色。

華語文口語測驗進階高階級直接對應的 CEFR B1 至 B2 等級口語指標中提及，此階段使用者可處理的主題與語言使用情境都以生活常見的、具體且個人感興趣，如一般工作、學校、休閒生活等主題的具體情境，以及和個人興趣、專業領域相關的熟悉、抽象主題、部分抽象的公共議題為範疇；ACTFL 大綱與其相近的指標內容則可見於部分 IH 及 AL、AM、AH 中。此間口語表達能力發展歷程亦高度相似，最顯著的整體特徵為「能直接、有組織性地描述」；歷程皆從「能較穩定掌握基本詞彙使用及句子語序，直接地敘述或描述」(B1↔IH~AL)發展至「敘述或描述時，結構清晰，基本用詞使用準確且多元豐富」(B1~B2↔AM)，甚而能「詳細說明、指出細節，使用例證、假說以佐證個人看法，組織連貫」(B2↔AH)；共通的表達及互動策略則包含有效地修正(B1~B2↔IH)或重新表述、重新措辭(B1~B2↔AL~AM)及迂回釋義(B1~B2↔AH)。

至於華語文口語測驗流利精通級直接對應的 CEFR C1 至 C2 等級口語指標中提及，此階段使用者在所有非正式場合及正式場合中，都能完成廣泛、複雜且抽象的專業領域、公共性、國際性等主題的表達或互動溝通任務，ACTFL 大綱與其相近的指標內容則可見於部分 AH 及 S、D 中；此間口語表達能力發展歷程，最顯著的共通特徵包含「能依聽者需求彈性調整說話方式」(C1↔D)、「能針對特定立場或觀點有效回應、闡述己見或他人論點」(C1~C2↔S~D)、「能採用目標語文化道地的表達方式，引用文化和歷史典故」(C2↔D)，組織結構的發展歷程也具備共通性，皆由「一定篇幅的連貫講述」(C1~C2↔AH)臻於「結構嚴謹、言簡意賅，接近書面語言」(C2↔D)，但仍可能在使用低頻句式及複雜的高頻句式時，出現零星但不影響聽者理

解，甚至可能無法察覺的語言錯誤（C1~C2↔S~D）。

此外，透過本研究結果，除了以 CEFR 或 ACTFL 大綱的分級來設計課程內容的教學者或研發語言測驗者，能清楚了解學習者或受測對象與 CEFR 或 ACTFL 大綱等級的對應情形，進而規劃後續語言學習的加強方向；也有助於受測者透過該測驗所參照的 CEFR 或 ACTFL 大綱分析自身語言能力，依個人需求擬定日後強化的語言能力向度；甚至對參考 CEFR 或 ACTFL 大綱的指標自學的學習者而言，都能利用熟悉的分級概念自行判斷學習成效，解讀測驗結果。綜言之，即是能在不侷限於受測者可能迥異的語言背景、目標語學習歷程的情況下，客觀地提供語言學習者或受測者其實際語言應用能力的資訊。

最後，由於華語文口語測驗尚未進行垂直等化，因此在跨等級相鄰級分的音檔比較與判斷不易操作。未來可進行垂直等化研究，建立完整的測驗級分對照關係，俾利臨界點的等級對應判斷結果更為精確。

## 引用文獻

- American Council on the Teaching of Foreign Languages. 2012. ACTFL Proficiency Guidelines. Accessed online, May 1, 2014. [http://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012\\_FINAL.pdf](http://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf)
- American Council on the Teaching of Foreign Languages. 2016. Assigning CEFR Ratings to ACTFL Assessments. Accessed online, May 1, 2016. [https://www.actfl.org/sites/default/files/reports/Assigning\\_CEFR\\_Ratings\\_To\\_ACTFL\\_Assessments.pdf](https://www.actfl.org/sites/default/files/reports/Assigning_CEFR_Ratings_To_ACTFL_Assessments.pdf)
- Bärenfänger, Olaf, and Erwin Tschirner. 2012. *Assessing Evidence of Validity of Assigning CEFR Ratings to the ACTFL Oral Proficiency Interview (OPI) and the Oral Proficiency Interview by computer (OPIc)*. White Plains, NY: Language Testing International. (Technical Report 2012-US-PUB-1). Leipzig: Institute for Test Research and Test Development.
- Cizek, Gregory J. 1993. Reconsidering standards and criteria. *Journal of Educational Measurement* 30.2: 93-106.
- Cizek, Gregory J. 2006. Standard setting. *Handbook of Test Development*, eds. by Thomas M. Haladyna, and Steven M. Downing, 225-258. Mahwah, NJ:

- Lawrence Erlbaum.
- Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge, UK: Press Syndicate of the University of Cambridge. Accessed online, January 17, 2007. [http://www.coe.int/t/dg4/linguistic/Source/Framework\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf)
- Council of Europe. 2009. *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment-A Manual [On-line]*. Accessed online, August 1, 2012. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680667a2d>
- Council of Europe. 2018. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (Companion Volume)*. Accessed online, February 16, 2018. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Jaeger, Richard M., Ina V. S. Mullis, Mary L. Bourque, and Sharif Shakrani. 1996. Setting performance standards for performance assessments: Some fundamental issues, current practice, and technical dilemmas. *Technical Issues in Large-Scale Performance Assessment*, ed. by Gary W. Phillips, 79-115. Washington, DC: National Center for Education Statistics.
- Jones, Neil, Karen Ashton, and Tamsin Walker. 2010. Asset languages: A case study of piloting the CEFR manual. *Aligning Tests with the CEFR: Reflections on Using the Council of Europe's Draft Manual (Studies in Language Testing) 33*, ed. by Waldemar Martyniuk, 227-248. Cambridge: Cambridge University Press.
- Kane, Michael. 1994. Validating the performance standards associated with passing scores. *Review of Educational Research* 64: 425-461.
- Language Testing International. 2016. Alignment of the ACTFL Proficiency Ratings to the Canadian Language Benchmarks (CLB). Accessed online, October 15, 2016. [https://www.languagetesting.com/pub/media/wysiwyg/research/white\\_papers/Alignment\\_of\\_the\\_ACTFL\\_Proficiency\\_Ratings\\_to\\_the\\_CLB\\_D7.pdf](https://www.languagetesting.com/pub/media/wysiwyg/research/white_papers/Alignment_of_the_ACTFL_Proficiency_Ratings_to_the_CLB_D7.pdf)
- Lewis, Daniel M., Howard C. Mitzel, and Donald R. Green. 1996. *Standard*

- Setting: A Bookmark Approach*. Paper presented at the Council of Chief State School Officers National Conference on Large Scale Assessment, June 2016. Phoenix, AZ.
- Martínez Baztán, Alfonso. 2008. *La evaluación oral: Una equivalencia entre las guidelines de ACTFL y algunas escalas del MCER*. Granada: Universidad de Granada doctoral thesis.
- Papageorgiou, Spiros, Richard J. Tannenbaum, Brent Bridgeman, and Yeonsuk Cho. 2015. *The Association between TOEFL iBT® Test Scores and the Common European Framework of Reference (CEFR) Levels* (Research Memorandum No. RM-15-06). Princeton, NJ: Educational Testing Service.
- Tannenbaum, Richard J., and E. Caroline Wylie. 2008. *Linking English-Language Test Scores Onto the Common European Framework of Reference: An Application of Standard-Setting Methodology*. TOEFL iBT Research Report, TOEFL iBT-06, RR-08-34. ETS, Princeton, NJ.
- Swender, Elvira, Erwin Tschirner, and Olaf Bärenfänger. 2012. Comparing ACTFL/ILR and CEFR based reading tests. *Aligning Frameworks of Reference in Language Testing: The ACTFL Proficiency Guidelines and the Common European Framework of Reference*, ed. by Erwin Tschirner, 123-138. Tübingen: Stauffenburg.
- Waldemar, Martyniuk. (ed.) 2010. *Aligning Tests with the CEFR*. Cambridge: Cambridge University Press.
- 王伯韜。2015。《英語專業四、八級考試與《歐洲語言共同參考框架》對接的可行性研究》。江蘇：揚州大學碩士論文。[Wang, Botao. 2015. *The Feasibility of Relating TEM to CEFR: Taking Reading Comprehension Test for Example*. Jiangsu: Yangzhou University MA thesis.]
- 王昕。2014。〈新 HSK 四級聽力測試效果與《歐洲語言共同參考框架》匹配度研究〉，《華西語文學刊》，第 1 期，193-201。[Wang, Xi. 2014. Comparing listening comprehension in new HSK-4 with Common European Framework. *Acta Linguistica et Litteraturaria Sinica Occidentalia* 1: 193-201.]
- 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 (3) 口語測驗信效度》。新北：國家華語測驗推動工作委員會。[Steering

- Committee for the Test Of Proficiency-Huayu. 2015. *Technical Report of TOCFL 2013-3: Reliability and Validity of the Speaking Test*. New Taipei: SC-TOP.]
- 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 2014 (3) 口語測驗信效度》。新北：國家華語測驗推動工作委員會。[Steering Committee for the Test Of Proficiency-Huayu. 2016. *Technical Report of TOCFL 2014-3: Reliability and Validity of the Speaking Test*. New Taipei: SC-TOP.]
- 國家華語測驗推動工作委員會。2017。《華語文能力測驗一百零六年成果報告》。教育部委託計畫案成果報告(未出版)。[Steering Committee for the Test Of Proficiency-Huayu. 2017. *TOCFL Project Outcomes Report 2017*. Ministry of Education commissioned project, unpublished.]
- 國家華語測驗推動工作委員會。2018。《華語文能力測驗技術報告 2016 (1) 口語測驗信效度》。新北：國家華語測驗推動工作委員會。[Steering Committee for the Test Of Proficiency-Huayu. 2018. *Technical Report of TOCFL 2016-1: Reliability and Validity of the Speaking Test*. New Taipei: SC-TOP.]
- 鹿士義。2011a。〈商務漢語考試 (BCT) 閱讀能力與歐盟框架的匹配研究〉，《語言文字應用》，第 1 期，81-90。[Lu, Shiyi. 2011. Comparing reading comprehension in Business Chinese Test (BCT) with Common European Framework (CEFR). *Applied Linguistics* 1: 81-90.]
- 鹿士義。2011b。〈商務漢語考試 (BCT) 與歐洲語言共同參考框架 (CEFR) 的等級標準關係研究〉，《華文教學與研究》，第 2 期，56-63。[Lu, Shiyi. 2011. The study on the corresponding scales between Business Chinese Test (BCT) and Common European Framework of Reference (CEFR). *TESOL Studies* 2: 56-63.]
- 黃婷、賈國棟。2012。〈語言測試與《歐洲語言共同參考框架》匹配的可行性研究〉，《外語測試與教學》，第 1 期，38-49。[Huang, Ting, and Guodong Jia. 2012. The feasibility study of linking language test to the CEFR—Taking College English Test as an example. *Foreign Language Testing and Teaching* 1: 38-49.]
- 趙家璧、藍珮君、陳柏熹。2016。〈CEFR 與 ACTFL 初級閱讀能力描述對應

- 研究：以「兒童華語文能力測驗」為媒介》，《臺灣華語教學研究》，第13期，41-64。[Chao, Chia-pi, Pei-jiun Lan, and Po-hsi Chen. 2016. Aligning ACTFL and CEFR proficiency levels: Utilizing children's Chinese competency certification. *Taiwan Journal of Chinese as a Second Language* 13: 41-64.]
- 劉靜觀。2012。《英語專業四級測試與歐洲語言共同參考框架的匹配研究》。河南：河南師範大學碩士論文。[Liu, Jingguan. 2012. *Aligning TEM-4 with The CEFR*. Henan: Henan Normal University MA thesis.]
- 藍珮君、陳柏熹、張可家、施泰亨、林玲英。2013。〈Yes/No Angoff 法在華語文口語測驗的應用〉，發表於「2013年第二屆標準本位評量國際研討會」，2013年11月14-15日。臺北：臺灣師範大學。[Lan, Pei-jiun, Po-hsi Chen, Ko-chia Chang, Tai-heng Shih, and Ling-ying Lin. 2013. Cut-off scores of the TOCFL Speaking test based on the Yes/No Angoff standard setting method. Paper presented at The 2<sup>nd</sup> International Conference on Standards-Based Assessment, November 14-15, 2013. Taipei: National Taiwan Normal University.]
- 羅蓮。2017。〈漢語分級測試與CEFR等級的連接研究〉，《語言文字應用》，第2期，110-118。[Luo, Lian. 2017. A study on the relations of the Chinese Placement Test and CEFR. *Applied Linguistics* 2: 110-118.]

[審查：2022.4.15 修改：2022.5.16 接受：2022.6.1]

廖才儀

Tsai-Yi LIAO

24449新北市林口區仁愛路一段2號 國家華語測驗推動工作委員會

Steering Committee for the Test Of Proficiency-Huayu

No.2, Sec. 1, Ren-ai Rd., Linkou Dist., New Taipei City 24449, Taiwan

Julis8814@gmail.com

藍珮君

Pei-Jiun LAN

24449新北市林口區仁愛路一段2號 國家華語測驗推動工作委員會  
Steering Committee for the Test Of Proficiency-Huayu  
No.2, Sec. 1, Ren-ai Rd., Linkou Dist., New Taipei City 24449, Taiwan  
martinalan@sc-top.org.tw

梁綺容

Chi-Jung LIANG

24449新北市林口區仁愛路一段2號 國家華語測驗推動工作委員會  
Steering Committee for the Test Of Proficiency-Huayu  
No.2, Sec. 1, Ren-ai Rd., Linkou Dist., New Taipei City 24449, Taiwan  
floraliang7@sc-top.org.tw

陳柏熹

Po-Hsi CHEN

10610臺北市和平東路一段162號 國立臺灣師範大學教育心理與輔導學系  
Department of Educational Psychology and Counseling  
National Taiwan Normal University  
No.162, Sec 1, Heping E. Rd., Taipei City 10610, Taiwan  
chenph@ntnu.edu.tw

## **Aligning the CEFR Oral Descriptors with the ACTFL Speaking Proficiency Guidelines Based on the TOCFL Speaking Test**

**Tsai-Yi LIAO Pei-Jiun LAN Chi-Jung LIANG**  
**Steering Committee for the Test Of Proficiency-Huayu**

**Po-Hsi CHEN**

**Department of Educational Psychology and Counseling, Chinese,  
Language and Technology Center  
National Taiwan Normal University**

### **Abstract**

This study aims to build an alignment between the Common European Framework of Reference (CEFR) and the ACTFL Proficiency Guidelines, utilizing the TOCFL Speaking test, which is designed based on the CEFR. 17 experts participated in a standard-setting process applying the Exemplar Response method which used the sample soundtracks for each test band and item type. The study provided procedural and internal validity to support the results of the proposed alignment between CEFR and the ACTFL Proficiency Guidelines.

The results showed that, in terms of speaking proficiency, CEFR level A1 is equivalent to ACTFL Novice High and Intermediate Low sublevels, while level A2 is equivalent to the Intermediate Low and Intermediate Mid sublevels. CEFR level B1 is lower than the ACTFL Intermediate High sublevel, while level B2 is equivalent to the Advanced Low and Advanced Mid sublevels. CEFR level C1 is equivalent to the ACTFL Advanced High sublevel and Superior level, while C2 is higher than the Superior level. The study revealed some shared characteristics between the CEFR and the ACTFL Guidelines which would be useful to guide further studies.

**Keywords:** the ACTFL Guidelines, the CEFR, Exemplar Response method, speaking test, standard-setting