

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

趙家璧、藍珮君、陳柏熹
國家華語測驗推動工作委員會

摘要

本研究採用國家華語測驗推動工作委員會參照 CEFR 初級使用者（A 等）能力指標所研發的兒童華語文閱讀測驗為媒介，透過書籤法（Bookmark Method）標準設定程序，連結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）的初級閱讀能力，期能分析 CEFR 與 ACTFL 初級閱讀能力描述的對應關係。本研究邀集 11 位專家採用書籤法進行兩回合的標準設定研究，使用兒童華語文閱讀測驗題目，對 ACTFL 初級與中級的能力描述進行標準設定。研究結果顯示：1. 低於 CEFR A1 以下的 Pre-A1 級約與 ACTFL 初級初等、初級中等相當，而 A1 級與初級高等、中級初等相當，至於 A2 級則與中級中等、中級高等相當；2. 在標準設定結果的效度驗證方面，本研究提出程序性效度與內部效度兩項資料；3. 在定義初級閱讀能力時，CEFR 與 ACTFL 有數項共同特徵和共通的學習進程。本研究結果有助於兒童測驗使用者透過 CEFR 與 ACTFL 的語言能力架構理解測驗結果，同時也提供華語文教學上有關 CEFR 與 ACTFL 對應的客觀依據。

關鍵詞：CEFR、ACTFL 能力指標、兒童華語文能力測驗、標準設定程序

1. 緒論

1.1 研究動機

標準化語言能力測驗設計的主要目的在為應試者提供公正、客觀且不受學習方式和歷程影響的語言程度評量。為達成這項任務，在編制語言能力測驗

時，採用具有公信力的語言能力描述架構作為測驗分級和試題設計的基礎，已成為語言能力測驗研製的共同趨勢。由於學習文化或指標應用範疇上的差異，各地區所制定的語言能力描述架構，無論是分級方式或各等級能力描述的內容，皆存在差異。因此，當標準化測驗採用特定架構為設計規準時，考試結果的參考價值就可能因不同的語言能力描述系統，而受到影響。

「歐洲共同語文參考架構」(The Common European Framework of Reference for Languages: Learning, Teaching, Assessment, 簡稱 CEFR) 與「ACTFL 語言能力指標」(ACTFL Proficiency Guidelines, 簡稱 ACTFL) 兩項語言能力描述架構，分別由歐洲理事會(Council of Europe)與美國外語教師協會(American Council on the Teaching of Foreign Languages)研發，是目前能見度最高、應用最廣的兩套語言能力描述架構，在世界各地的語言教學、語言測驗等領域廣受採用。雖然這兩個語言能力指標在分級尺度與能力描述上有所出入，但這兩個架構的基礎都是以真實語言情境和達成溝通任務做為語言能力的評估依據。因此，無論從語言教學、語言測驗或是學習者自評的角度來看，歐、美兩套語言能力架構所描述、分級的語言能力應能互相對照、比較。

本研究利用國家華語測驗推動工作委員會(簡稱華測會)所研發之標準化語言能力測驗「兒童華語文能力測驗」(Children's Chinese Competency Certification, 以下簡稱兒童測驗或 CCCC)為媒介，以比較 CEFR 與 ACTFL 的語言能力描述，一方面能幫助兒童測驗的使用者透過自身熟悉的語言能力架構理解測驗結果，一方面也希望透過質化與量化並重的研究方式，為使用 CEFR 與 ACTFL 兩套語言描述的語言教學者或語言學習者提供更為客觀且精準的對應參考。

1.2 研究問題

本研究希望由質性與量化的角度分析以下兩個問題：

1. CEFR 定義下的初級使用者閱讀能力與 ACTFL 能力指標分級的對應關係為何？
2. CEFR 與 ACTFL 兩個能力指標在不同的分級方式下，對二語學習者初級的閱讀能力、學習進程與語言表現方面有哪些共同特徵？

2. 文獻探討

2.1 CEFR 與 ACTFL 語言能力指標

CEFR 和 ACTFL 是目前能見度最高、應用較廣的能力描述，CEFR 語言能力指標的制定者為歐洲理事會，其主要的編制目的是考量歐洲各國的教育體系

與課程的安排迥異，希望為境內語言教學相關領域之專業人士克服溝通障礙、提供共同的語言（Council of Europe 2001）。而 ACTFL 語言能力指標的制訂與施行則是以美國境內的需求為考量，ACTFL 語言能力指標最初改寫自美國聯邦組織「政府跨機構語言圓桌」（U.S. Government's Interagency Language Roundtable）所編制的語言量表（ILR Skill Level Descriptions），目的是為美國境內學術社群提供更適用的語言能力指標。歷年來為反應實際教學、測驗的需求，曾經過三次修訂，而本研究所針對的 2012 年增修版中，即表明此語言指標旨在於全球環境下為職場與學術場域提供語言能力的標準（American Council on the Teaching of Foreign Languages 2012a）。

這兩個語言能力指標都以語言實際溝通任務做為能力描述的基礎，主要的編制目的都是為語言學習相關的社群提供共通的專業語言。雖然這兩個能力指標都以類似的專業社群為目標使用者，制訂指標的方向與理論也十分雷同。但由於制訂機構對等級有不同的定義，以及描寫能力內容的差異，表面上，CEFR 與 ACTFL 等級間的對應關係並不明顯。

本小節將簡介兩套指標個別的編制原則與分級方式，再簡單討論過去研究在兩項能力指標對應上的缺乏，以及本研究針對初級階段學習者閱讀能力描述之對應採取的基本分析方向。

2.1.1 CEFR 能力描述之初級使用者

歐洲的 CEFR 能力指標以行動導向與社會導向為編制原則，不以特定的教學方式或教材內容為基礎，主要的目的是希望為語言教學者與學習者提供語言間互相認可的客觀標準。CEFR 將語言學習者需習得的語言能力分為「初級使用者」（A 等）、「獨立使用者」（B 等）與「精通使用者」（C 等）三等，每等之下再各自分出兩個次等級，共三等六級。在聽力、閱讀、口頭表達、書寫表達幾項能力下再依照不同語言活動的目的提供分項描述（Council of Europe 2001）。

本研究所針對的「初級使用者」閱讀能力，分為 A1 與 A2 兩級，其中整體閱讀能力與根據語言活動不同所定義的分項能力敘述如表 1 所示。由描述中可了解本階段的語言使用者的閱讀能力限於具體、熟悉並與個人生活最切身相關的情境和主題，僅能處理簡短、簡單的文本中的要點，並在理解文本時多仰賴對基本、常見的詞彙的理解。

表 1 CEFR A1 至 B1 閱讀能力描述*

等級 向度	A1	A2	B1
整體性 閱讀理解	能理解非常簡短、簡易的文本，但一次只讀一個片語，挑出熟悉的名字、詞語與基本片語，需要常常地重新閱讀。	1. 能理解含常用字彙的短文，包括相當比例國際上的共用的字彙。 2. 對於具體、熟悉的主題，能清楚理解其簡短、簡單的內容，這種主題含日常或與工作有關的常用語言。	能以令人滿意的理解程度，來閱讀很直接表達的事件的文本，而這些事件是他感興趣的主題。
閱讀書信	能理解明信片上簡短、簡單的訊息。	1. 能理解簡短、簡單的個人訊息。 2. 針對熟悉的主題，能理解標準慣用的書信及傳真的基本形式（詢問、訂購、確認書信等等）。	在私人信件中能完全理解事件、情感及願望的描述，以至可以與筆友作規律的書信往返。
導向閱讀	在最平凡的日常情境中，能辨識出簡單告示中的熟悉的名字、詞語及非常基本的片語。	能在日常生活簡易材料（廣告、旅遊指南、菜單、參考表及時間表）中找出特定可預知的訊息；能在清單中找到特定訊息，並區隔所要的資訊（如使用「黃頁」來尋找服務項目或商家）。能理解在公共場合（街道、餐廳、鐵路車站）或工作場所（說明、指導、危險警告）日常的標示及告示。	1. 能掃描較長篇的內容，以便尋找想要的資訊，並且從內容中不同的部分或不同的內容中蒐集資訊以便完成一個特定任務。 2. 在日常的材料，例如：信件、小手冊及簡短官方文件中，能找到並理解相關資訊。
為資訊與 論證閱讀	在有視覺方面的協助時，對於較簡單、提供資訊類的簡短敘述，能取得敘述內容中的看法。	能辨認出較簡單書寫材料的特定訊息，如：書信、小手冊及描述事件的報紙。	1. 在清楚標示的論證內容中，能辨認主要結論。儘管不一定會很像細對於某些議題的討論，仍能辨認出論證方式。 2. 對於熟悉的主題且以直接手法呈現的報紙文章中，能辨認出重點。
閱讀說明 指導	能跟上簡短、簡單的書寫說明（例如：從甲地到乙地）。	1. 對於日常生活中遭遇的設備，例如：公共電話，能了解簡單的說明。 2. 能了解以簡單語言的表達，如安全方面的規定。	對於一種設備，能了解書寫清楚，直接的說明。

* 莊永山等（2007:67-69）

2.1.2 ACTFL 能力描述

美國的 ACTFL 語言能力指標於 1986 年首次公布，第一版的內容包括對聽力能力與閱讀能力的描述，期間再於 1999 和 2001 年修訂寫作能力和口語能力的指標描述。2012 年公布的最新修訂版，除針對既有的能力分級增加一般性說明外，再細分次等級，並增加口語能力與寫作能力更高等級的優異級能力描述。ACTFL 能力指標的制訂原則，不以任何教學理論、教學法或語言課程為本，而是根據未經事先準備的真實語言情境，來描述語言學習者聽、說、讀、寫等分級能力的內容。綱要中所描述的能力由低至高共分為初級（Novice），中級（Intermediate），高級（Advanced），優秀（Superior）和優異（Distinguished）五級，如表 2 所示，本架構在能力較低的三個等級中又再細分為：初等（Low）、中等（Mid）與高等（High）等次等級（ACTFL Proficiency Guidelines 2012a）。本研究以初級閱讀能力為主，在閱讀能力描述中，與初級學習者較有相關的是初級（Novice）與中級（Intermediate）及其次等級的能力描述。表 3 是 ACTFL 有關 ACTFL 初級與中級閱讀能力的描述。

表 2 ACTFL 能力描述架構

優異（Distinguished）	
優秀（Superior）	
高級（Advanced）	高等（High）
	中等（Mid）
	初等（Low）
中級 （Intermediate）	高等（High）
	中等（Mid）
	初等（Low）
初級 （Novice）	高等（High）
	中等（Mid）
	初等（Low）

（ACTFL Proficiency Guidelines 2012a）

表 3 ACTFL 初級與中級閱讀能力描述

初級（Novice）
初級水準的讀者能夠理解關鍵字和同源詞，以及與語境高度一致的模式化短語。
初級水準的讀者能夠從對話題或語境非常熟悉的具高預測性的文本中獲得有限數量的資訊，例如旅館帳單、信用卡收據或天氣圖。初級水準的讀者主要依賴他們自己的背景知識和語言以外的輔助（例如天氣圖上的圖像或信用卡帳單的格式）來獲取含義。
初級水準的讀者在能夠預測文本資訊時，可以最好地理解文本。在初級水準，對關鍵字、同源詞及模式化短語的識別使得理解成為可能。

表 3 ACTFL 初級與中級閱讀能力描述（續）

初級（Novice）	
初等 （Low）	初級初等水準的讀者能夠識別有限的幾個字母、符號或字符。他們偶爾能夠在有良好語境的支持下識別高頻詞及/或短語。
中等 （Mid）	初級中等水準的讀者能夠識別字母表或音節書寫系統的字母或符號，或是字元語言中的有限幾個字元。他們能夠識別若干個與語境高度一致的詞和短語，包括同源詞和外來詞，但極少能夠理解超出單個短語的材料。經常需要重複閱讀。
高等 （High）	初級高等水準的讀者能夠完整且比較輕鬆地理解一系列與語境高度一致的文本中的關鍵字和同源詞以及模式化短語。如果學過詞彙，他們可以讀懂可預測的語言和訊息，例如列車時刻表、地圖及路牌上的語言。初級高等水準的讀者通常能夠借助語境或語言之外的輔助，從傳達基本資訊的簡短且不複雜的文本中獲取其含義。
中級（Intermediate）	
在中級水準上，讀者能夠理解簡單、可預測、連接鬆散的文本所傳達的資訊。讀者主要依賴語境線索。如果是熟悉的文字模式，例如天氣預報或社會聲明，他們最容易理解資訊。 中級水準的讀者能夠讀懂傳達基本資訊的文本，例如在聲明、通知、線上公告板和論壇中包含的文本。這些文本並不複雜，而且具有可預測的呈現模式。話語間的連接很少，基本由單個句子及含有大量高頻詞彙的一連串句組構成。 當從簡單、直截了當的文本中獲取含義時，中級水準讀者的理解最為準確。他們能夠理解高度熟悉的日常語境中所含的訊息。這個水準的讀者可能無法完全理解充滿細節的文本，或是那些需要具備語言結構知識以理解次序、時間框架及時間順序的文本。	
初等 （Low）	中級初等水準的讀者能夠從涉及有限的個人及社會需求的最簡單的連貫文本中理解部分資訊，不過閱讀中可能會頻頻出現誤解。這個水準的讀者在理解任何長度的連貫文本的含義時都會有困難。
中等 （Mid）	中級中等水準的讀者能夠理解傳達基本資訊的簡短且不複雜的和涉及個人興趣或知識的基本個人及社會話題的文本，不過可能出現一些誤解。如果是熟悉的話題，這個水準的讀者或許能從帶有描述及敘述的簡短連貫文本中獲取一定含義。
高等 （High）	中級高等水準的讀者能夠完整輕鬆地理解傳達基本資訊的簡短且不複雜的文本，憑借個人興趣或知識，應對個人及社會話題。這一水準的讀者還能讀懂一部分帶有描述及敘述的連貫文本，不過由於其在詞彙、結構以及語言書寫規範方面的知識有限，理解上會偶爾出現空白。

（ACTFL Proficiency Guidelines 2012b）

2.1.3 CEFR 與 ACTFL 能力描述之對應

在語言教學相關領域中，基於對語言教學的課程安排與解釋語言能力測驗結果的實用需求，比較不同語言能力指標的對應情形，是一個值得討論的議題。過去研究 CEFR 與 ACTFL 這兩個能力指標間的對應方式大致可分為兩類，一類是試圖透過分析能力指標中所描述的語言功能與活動，比較兩個指標

描述內容的異同，並連結其語言等級；另一類則以測驗結果或語言學習者產出的語料為媒介。

在 CEFR 與 ACTFL 兩個指標對應的研究中，多以產出型的語言能力，尤其是學習者的口語樣本為媒介，研究中評分者以 CEFR 與 ACTFL 的能力指標為同一群受測者的口語樣本分級，研究結果顯示 ACTFL 的初級（Novice）至中級（Intermediate）的語言能力與 CEFR 的 A 等至 B1 之間的能力相當（Martínez Baztán 2008; Bärenfänger and Tschirner 2012）。雖然本研究主要探討屬於接收型的閱讀能力，但透過口語能力的研究結果，對了解兩項指標分級的原則仍有所幫助。

根據表 1 與表 3 所述，華測會四名擔任閱讀測驗試題研發的研究人員在執行標準設定前，先召開會議討論 CEFR 與 ACTFL 的閱讀能力指標，並比對各等級所描述的語言能力特徵。ACTFL 初級下的三個次等級，對於使用者的能力描述都著重在理解常見、高頻詞語的能力，且強調該等級語言使用者對文本的理解需在詞語與文本情境間具高度相關性且具可預測性的情況下方能達成。其中理解日常生活常見、高頻詞語，以及需情境輔助等特徵，都與 CEFR 對 A1 等級閱讀能力的描述相似。

ACTFL 中級的能力描述，著重於語言使用者在與個人生活、社會情境下的閱讀能力。相較於 ACTFL 初級，ACTFL 中級的使用者已從辨認詞語、短句階段過渡到對簡短、連貫敘事文本的理解；但在 ACTFL 中級初等的描述中，閱讀者能理解「部分資訊」的文本仍受限於「有限的個人及社會需求」的範圍內，較接近 CEFR A2「整體性閱讀理解」中所界定的文本類型，因此推估大致與 CEFR A2 的能力相近。ACTFL 中級中等描述中界定的「涉及個人興趣」與「熟悉的話題」等閱讀內容介於 CEFR「整體性閱讀理解」中 A2 至 B1 等級中所定義的文本，而 ACTFL 定義中級高等的使用者能「完整輕鬆地理解傳達基本資訊的簡短且不複雜的文本……」的描述則與 CEFR B1 等級所描述的「能以令人滿意的理解程度，來閱讀很直接表達的事件的文本……」的閱讀能力相似。因此，研究人員通過上述討論，初步判斷 CEFR 的 A 等至 B1 的閱讀能力描述大致與 ACTFL 架構中的初級至中級的描述相似。

在綜合考量前人針對口語能力的研究結果，與華測會試題研發人員分析、比對的結論，本研究接下來針對 ACTFL 與 CEFR 之 A 等閱讀能力的對應研究，將聚焦在 ACTFL 初級至中級的能力範圍之內。

2.2 兒童華語文能力測驗與 CEFR 等級對應

2.2.1 兒童測驗簡介

本研究以「國家華語測驗推動工作委員會」所研發的兒童華語測驗為對應媒介。該測驗根據 CEFR 初級語言使用者（basic user）的能力描述，針對七至十二歲母語非華語之學習者而設計的標準化語言能力測驗，2008 年研發完成進行預試，2009 開始舉行正式考試，由易至難分為萌芽級、成長級與茁壯級三等。

其中成長級與茁壯級分別依據 CEFR A1 與 A2 的能力描述進行規劃與設計，測驗難度與同樣參照 CEFR 進行研發的華語文能力測驗（Test of Chinese as a Foreign Language，簡稱 TOCFL）入門級、基礎級相當；同時，針對初學華語，且年齡較小的學習者，基於鼓勵學習成就與建立學習自信的出發點，兒童測驗自 A1 級能力所描述之簡單語言任務中，往下針對更為基礎的能力研發出萌芽級，並將萌芽級定為 Pre-A1 級（陳怡靜、趙家璧 2012）。表 4 為兒童華語文能力測驗之能力描述（國家華語測驗推動工作委員會 2015a）。

表 4 兒童華語文能力測驗能力描述

測驗等級	閱讀理解能力
萌芽級	能理解基本的詞語及簡單的句子。在視覺輔助的前提下，藉由掌握基本詞語能大致理解簡單的書寫材料。
成長級	在視覺輔助下，能理解簡單敘事短文的大意及找出重要訊息。能從日常生活的簡易書寫材料，例如菜單、車票、明信片及標示中，讀出基本訊息，例如：姓名、日期、時間、價錢、地點等。
茁壯級	能理解主題具體、與個人生活經驗相關的簡單敘事短文。能辨識日常生活中不同書寫材料的功能，且能從書寫材料，例如留言、便條、信件、學校公告、廣告或海報中，讀出重要的訊息，例如：活動時間、地點、辦法、適用對象、注意事項等。

兒童測驗作為標準化語言能力測驗不以任何特定教材為命題依據，測驗題材取自真實的生活情境，內容涵蓋所有與兒童經驗相關之主題，如日常生活與休閒活動、交通運輸、人物特徵、身體與健康、學校用語、時間與空間概念以及氣候與季節等。測驗實施方式為紙筆測驗，聽力測驗與閱讀測驗合併施測。

兒童閱讀測驗根據上述設計理念共包含看圖辨義、圖文連連看、閱讀材料與短文理解四種題型，各題型之所測能力與作答方式請參見表 5（國家華語測驗推動工作委員會 2014）。

表 5 兒童華語文閱讀測驗題型簡介

題型	測驗目標與使用等級	作答方式
看圖辨義（是非）	能理解常見、簡單的詞彙。 （萌芽級）	考生會看到一張情境圖片與二至三個簡單的句子。考生要判斷句子的意思是否和圖片內容相同。
圖文連連看	能理解簡單、簡短的句子。 （萌芽級、成長級）	考生會看到四個句子與五張圖片，考生需將句子連到正確的圖片，其中有一張圖片不會用到。
閱讀材料（選擇）	能理解簡短、簡單、書寫清楚的信件、告示、或日常生活中常見材料的基本形式、要點與可預知的訊息，如：明信片、時間表、廣告、菜單等。 （萌芽級、成長級、茁壯級）	考生會看到一則書寫材料，每則材料有一至三個題目，考生根據所讀的材料內容，選出正確答案。
短文理解（是非）	能理解簡單、簡短的敘事文，掌握文中主要情節和要點。 （成長級、茁壯級）	考生會看到一段敘事短文，每段短文配合三至四個句子，考生根據短文的内容判斷句子是否正確。

2.2.2 兒童測驗通過門檻制訂

在 2009 年兒童測驗正式推出時，萌芽級、成長級、茁壯級的題庫採取各等級內水平等化。2013 年，華測會為提高測驗通過門檻的精準度，對兒童測驗題庫進行兩項等化程序：1. 萌芽、成長、茁壯三等測驗題庫難度之垂直等化，使兒童測驗不同等級間的試題難度可相互比較；以及 2. 兒童測驗與 TOCFL 試題難度參數量尺的等化，使兒童測驗試題難度可與 TOCFL 相互參照。

第一項程序將兒童測驗三等題庫等化的方式是由萌芽級、成長級、茁壯級三等測驗題庫中分別挑出難、中、易不同難度區間之試題，組成聽力、閱讀測驗各一卷等化用試卷，聽力、閱讀測驗題數均為 40 題，施測樣本分別為聽力測驗 1,244 人次和閱讀測驗 1,231 人次。原先各自獨立的萌芽級、成長級與茁壯級測驗題庫，藉由此等化卷中試題難度的估計結果，進一步估計出題庫內其他試題的難度參數，讓各等級測驗的試題難度參數可連結至同一把量尺之上，不同等級間的試題難度也自此得以相互比較（國家華語測驗推動工作委員會 2015a）。

第二項等化程序是將兒童測驗難度參數連結至 TOCFL 之難度參數量尺，TOCFL 包含入門基礎級（A 等）、進階高階級（B 等）、與流利精通級（C 等）三等測驗，已於 2013 年透過 Angoff 法（詳下文）進行標準設定研究，並依此訂定出完整對應至 CEFR 三等六級之通過門檻（國家華語測驗推動工作委

員會 2015b)。為將兒童測驗試題難度連結至 TOCFL 試題難度參數量尺，以依循 TOCFL 標準設定的通過門檻結果，華測會邀集受測者同時作答前述之兒童測驗等化用試卷與 TOCFL 入門基礎級測驗的等化用卷。TOCFL 等化用卷之聽力、閱讀測驗題數均為 30 題，施測樣本分別為聽力測驗 183 人次和閱讀測驗 191 人次。固定入門基礎級已知試題難度參數後，所估計出的兒童測驗試題難度參數便能連結至 TOCFL 難度參數量尺。等化程序完成後，同為 CEFR A 等測驗之兒童測驗與入門基礎級測驗間，試題難度得以相互比較，其中成長級與茁壯級的通過門檻分別與 TOCFL 入門級（A1）和基礎級（A2）相同，而萌芽級（Pre-A1）通過門檻則由研發人員會同測驗專家參考茁壯級與成長級門檻差距後共同訂出（國家華語測驗推動工作委員會 2015a）。因此，兒童測驗的成長級與茁壯級的通過門檻分別對應至 CEFR 的 A1 和 A2 等級，而萌芽級的通過門檻是低於 CEFR A1 的語言能力。

2.3 標準設定方法與選擇

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），均採用標準設定程序進行測驗與語言能力架構的對應（Tannenbaum and Wylie 2008）。因此本研究希望邀集相關領域專家，透過標準設定程序以了解 CEFR 與 ACTFL 兩套架構間的對應關係，標準設定的研究方法十分多元，其中以 Angoff 法與書籤法（Bookmark Method）最為常見，以下將簡介這兩種標準設定方法、說明本研究對標準設定方法的選擇，以及標準設定實施的程序。

2.3.1 Angoff 法

William Angoff（1971）提出的 Angoff 法是標準設定研究中常見的方法之一，而由原始實施步驟變化產生的衍生研究方法也十分多樣。基本的 Angoff 法要求標準設定會議的參與者審閱題本中的每道試題，判定某等級或能力描述下之「最低能力」受測者答對每一道試題的機率；接著根據自己的標記結果與其他參與者說明、討論，並根據討論結果進行下一輪答對率的標註，最後再求得參與者對試題難度判定的平均值做為設定標準。由於原始的 Angoff 法要求參與者逐題決定最低能力者答對的機率，在題本組成之試題數量較多或需判斷的能力標準較多的狀況下，應用此方法的決策過程過於繁複，不但將使標準設定會議變得冗長，參與者間的評定結果也較難達到相當的一致性。有鑑於此，Angoff 法衍生出數種簡化參與者難度判定過程的研究方法，如改良式選擇型 Angoff 法（Modified Multiple Choice Angoff）（Berk 1986）與 Impara and Plake（1997）所提出的 Yes/No Angoff 法等。以此兩種衍生研究方法為例，選擇型

Angoff 法顧名思義，提供參與者 5%、20%、40%、60%、75%、90%、95% 等固定答對率，做為判定選項，而 Yes/No Angoff 法更進一步簡化判定程序，參與者僅需逐題判斷最低能力者能否答對該題即可。

2.3.2 書籤法 (Bookmark Method)

書籤法 (Lewis, Miztel and Green 1996) 與 Angoff 法同為目前十分常見的標準設定研究方法，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，為測驗訂定通過標準的方法。而書籤法與 Angoff 系列研究方法間的區別在於檢視題本內容進行難度判定的程序；Angoff 法要求標準設定的參與者逐題檢視標準設定題本中的每一道試題，並判斷最低能力者答對各題的可能性。書籤法則在檢視題本內容的步驟採用較簡易的執行方式，對能力描述等級或測驗通過標準的判定，參與者僅需在一系列依難度排列的題目中，指出最低能力者具有特定答對率的單一試題即可。此外，Angoff 法不一定會使用試題反應理論 (Item Response Theory, 簡稱 IRT) 的參數，但由於書籤法在實施上需將不同測驗題本中的試題依照難度置於同一量尺上排序，因此常採用 IRT 參數表示試題難度。

由於此次研究參與者須共同判斷 ACTFL 語言能力描述中六個次等級的難度標準，使用 Angoff 法比較語言等級間的異同，有兩個缺點：一、程序耗時；二、過去相關研究相當有限，事前無法決定各等級題本含括 ACTFL 哪幾個等級，施行上較為困難。同時，兒童測驗雖分為三個等級，但過去已透過 IRT 等化技術將不同等級試題難度串連在一起 (國家華語測驗推動工作委員會 2016)，加上組成題本的兒童華語文閱讀測驗試題以是非題與選擇題為主，融合 IRT 的書籤法較 Angoff 法更適合本研究的實施；故決定採用書籤法做為本研究的主要研究方法。

3. 研究方法

實施書籤法包括執行標準設定會議前的題本組成工作、邀請專家舉行標準設定會議，以及會議後分析標準設定結果三個階段。本節將先簡介本次標準設定使用的題本組成、會議參與者以及等級標記方式，再逐步詳述本研究利用書籤法執行標準設定的流程。

3.1 OIB 組成

首先需將一系列試題由易至難組成測驗卷 (Ordered Item Booklets, 簡稱為 OIB)，組成本次書籤法研究的 OIB 試題，皆挑選自兒童華語文閱讀測驗題庫，試題都經由 IRT 之 Rasch 模式得到難度參數 (Rasch 1960)，並由易到難

排序；組成 OIB 題本時參考萌芽級、成長級、茁壯級測驗正式考試所採用的試題難度比例以及各難度區間於題庫中所佔的比例。此外，考量到兒童閱讀測驗的四個題型中，因「圖文連連看」採用題組型配合題的作答方式，在型式上較不適合書籤法每頁僅呈現一道試題的特性，因此本研究不選用「圖文連連看」題型的試題；綜合考量以上的條件後，本研究共自兒童閱讀測驗中選出 85 題，組成得以代表兒童華語文閱讀測驗題庫現狀之 OIB。

3.2 操作方法

本研究希望透過書籤法判定 ACTFL 閱讀能力描述自初級至中級分級下各次等級之受測者能力值，由於 ACTFL 閱讀能力描述的分級十分精細，主要分級加上次等級共五級十一等，在研究中難從某次等級的描述中再定義「最低能力者」表現，為確保與會專家對判定標準能掌握高度一致性，在與測驗專家討論研究程序後，決定以各次等級典型的能力描述做為書籤法中判定 OIB 各等落點的標準。因此，與會專家在本研究中需判定符合 ACTFL 能力描述中初級及中級下六個次等級典型能力的使用者，在 OIB 中具有 67%答對機率的題目；最後再以相鄰兩等級的平均值做為較高等級的最低能力值，也就是切截點（cut off points）。

3.3 參與者

本次標準設定會議於 2014 年 8 月舉行，邀請的參與者皆為資深華語教師以及語言教學領域學者，共 11 人（表 6）。其中為資深華語教師身分的參與者皆擁有任教於中華民國境內華語文教學中心或海外大學華語課程的經驗，最資淺者有 5 年以上經驗，平均具 10 年以上華語教學年資，教學對象包括兒童、青少年與成人學習者，其中最資深者具 30 年以上教學經驗，並曾分別擔任華語師資培訓講師、初級華語或少兒華語教材編撰與教學法設計等相關經歷。而與會的語言教學領域學者則為研究語言學、漢語語法、華語教學等相關領域的專家。

表 6 研究參與者背景分布表

		人數	比例
性別	女	11	100.0%
	男	0	0.0%
專業背景	資深華語教師	8	72.7%
	語言教學領域學者	3	27.3%

3.4 研究步驟

- (1) 於會議召開兩週前提供相關資料給與會者先行研讀，包括 ACTFL 閱讀能力指標、本研究動機、研究方法與議程的介紹，以熟悉本研究所使用之能力指標與標準設定流程。
- (2) 會議當天先簡介標準設定的目的、書籤法原理、題本組成方式、標準設定程序，以及將進行標記的等級數量。
- (3) 說明 ACTFL 閱讀能力指標的分級方式與能力描述，並比較、討論相鄰等級或次等級間所描述的能力差異。
- (4) 介紹 OIB 中的兒童測驗題型種類，包括各題型設計方向與受測者應試時在題本中的呈現與作答方式。
- (5) 與會者於一份 85 題的閱讀測驗 OIB 中開始第一輪的標記，並寫下判定原因。其中程度最低與最高的初級初等與中級高等兩個等級，若與會者認為 OIB 中沒有符合這兩項等級描述的書籤標記點，則可選擇不標記。
- (6) 提供與會者根據步驟（5）的判斷結果所得之回饋訊息，包括：（a）每位參與者第一輪的標記題號；（b）透過判斷結果和試題難度參數，得出的能力值描述統計數值¹，包括平均數、中位數、最小值、最大值與四分位差；（c）OIB 試題之難度參數。接著，與會者根據上述回饋資訊，依等級順序由初級初等開始進行討論。
- (7) 完成第一輪討論後，與會者進行第二輪的標記，判斷方式同步驟（5）。
- (8) 根據步驟（7）之第二輪判斷結果，提供與會者如步驟（6）之回饋訊息，並進行第二輪討論。
- (9) 標記與討論的程序設定在各等級判定能力值之四分位差降至 0.5 以下後停止，因本次標準設定的各等級能力值之四分位差在第二輪皆降至該範圍內，故設定程序於第二輪結束。
- (10) 發放問卷調查與會者對本次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，研究者以 Rasch 模式答對率為 67%計算出十一名研究參與者於每一輪 OIB 初級初等、初級中等、初級高等、中級初等、中級中等，以及中級高等

¹ 若參與者於 OIB 未標記初級初等或中級高等的試題，為利於計算，分別以 OIB 第一題（最容易）與最後一題（最困難）的試題換算出能力值。

共六個標記點（即試題），每道已知難度參數試題所相對應之典型語言使用者能力值。然後以 SPSS 21 計算第一輪與第二輪參與者判定各等級典型使用者能力值描述統計結果，包括中位數和四分位差。因為平均數和標準差較容易受到極端值影響，而中位數與四分位差受的影響較小。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。四分位差為將所有數字依照由小到大依序排列後，排列在第三個四分位數（Q3）減去第一個四分位數（Q1）後的一半，數字越小表示數值分配情形的集中程度越高。

接著，再根據得到之六個等級第二輪的典型使用者能力值中位數，計算各個等級通過門檻的能力值，計算方式為相鄰兩等級典型能力值的平均數。如：初級初等與初級中等的典型能力值的平均數即為初級中等的最低能力值，也就是該等級的門檻，其餘等級以此類推。但因初級初等為最低等級，該等級通過門檻的計算方式調整為，先將標記為初級初等試題以答對率 33%換算出能力值並求得中位數後，再與初級初等典型能力值求中間值。

最後，分析問卷調查結果，計算參與者在各題的平均同意程度，若勾選「非常同意」則標記為 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，藉此了解參與者對於整個標準設定流程的理解程度。

4. 研究結果

由於 2 號參與者第一輪與第二輪在六個等級中僅判斷四個等級，且八筆標記中，有七題換算後之能力值皆為所有參與者中最低的，也和平均值差異相當大。顯示其判斷結果屬於極端值，研究者判斷此筆資料會對結果分析造成影響，故刪除其標記資料，以下結果皆從十位與會者的標記資料分析而來。

4.1 標準設定結果

表 7 為第一輪與第二輪參與者判定各等級典型能力值的描述統計結果，由表 7 可知，經過第二輪的判定後，除了初級中等和中級高等外，其餘四個等級的能力值中位數都較第一輪略微提高，且四分位差也都在第二輪降至 0.3 以下。兩輪之間四分位差變化最大的等級為中級高等，而該等級在第二輪的四分位差則成為六級間最低的。整體而言，與會專家在第一輪與第二輪對各等級的能力值判斷變化不大，四分位差均低於 0.5，代表本研究參與者對於各等的語言能力有十分一致的理解與判定標準。

表 7 第一輪與第二輪參與者判定各等級典型能力值描述統計結果

等級	第一輪		第二輪	
	中位數	四分位差	中位數	四分位差
初級初等	-4.275	0.164	-4.200	0.164
初級中等	-3.538	0.279	-3.538	0.103
初級高等	-2.921	0.244	-2.740	0.228
中級初等	-2.365	0.378	-2.097	0.280
中級中等	-1.407	0.289	-1.399	0.161
中級高等	-0.667	0.466	-0.667	0.038

兩輪的詳細判斷變化詳見圖 1 和圖 2，從圖 1 可知，在第一輪標記中，2 號與 10 號參與者認定具初級初等能力的受測者無法答對 OIB 題本中任何題目；7 號參與者則認定具中級高等能力的受測者能答對 OIB 題本中所有題目。此外，第一輪標記結果，相鄰等級間判定位置重疊的情形較明顯，以初級高等、中級初等與中級中等三個等級的第一輪標記為例，可發現多位參與者對相鄰兩等的標記落在相同的或緊鄰的題號之上。從圖 2 可觀察到，參與者經過對第一輪判定的表述與討論後，在初級高等與中級初等間以及中級初等與中級中等間，標記重疊度高的情形也明顯在第二輪中得到改善，第二輪判定的結果也在各等級中都較第一輪更趨一致。

同時，從圖 1 和圖 2 中可看出 2 號參與者在標記上的特殊性，第一輪中，2 號參與者未標記能力最低的初級初等與初級中等兩個等級，即使經過第一輪後的討論，2 號參與者在第二輪判定時仍選擇不標記這兩個等級。並且這位參與者兩輪的標記點，除第一輪對中級高等的判定外，所有等級判定的能力點都是與會者中最低的，例如第二輪判定時，十名參與者對中級初等的標記分布於第 41 題至第 55 題間，2 號參與者的標記點則在更低的 37 題上。如此一貫偏低的標記，也使這位參與者的判定結果成為極端值，於本研究中不予採計。

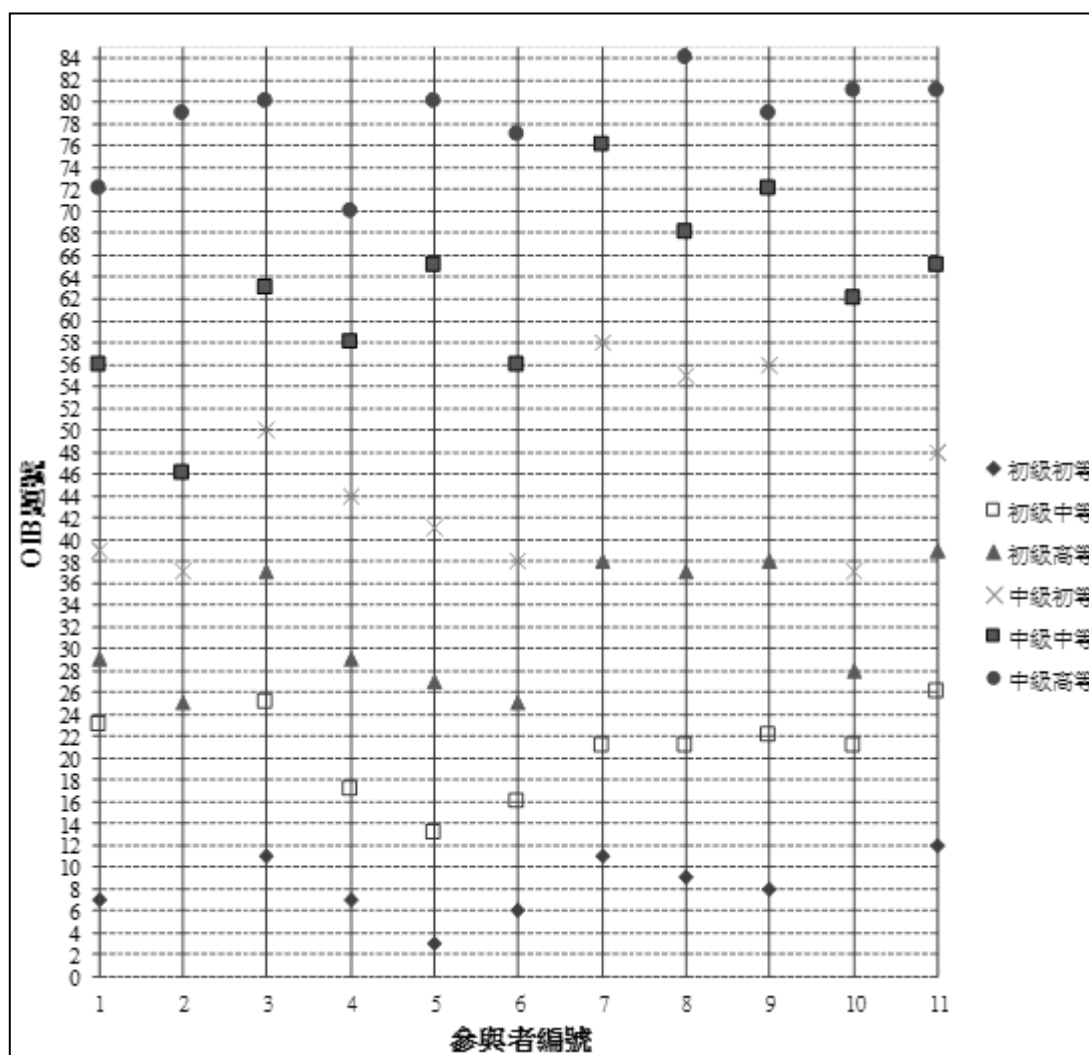


圖 1 第一輪標記結果

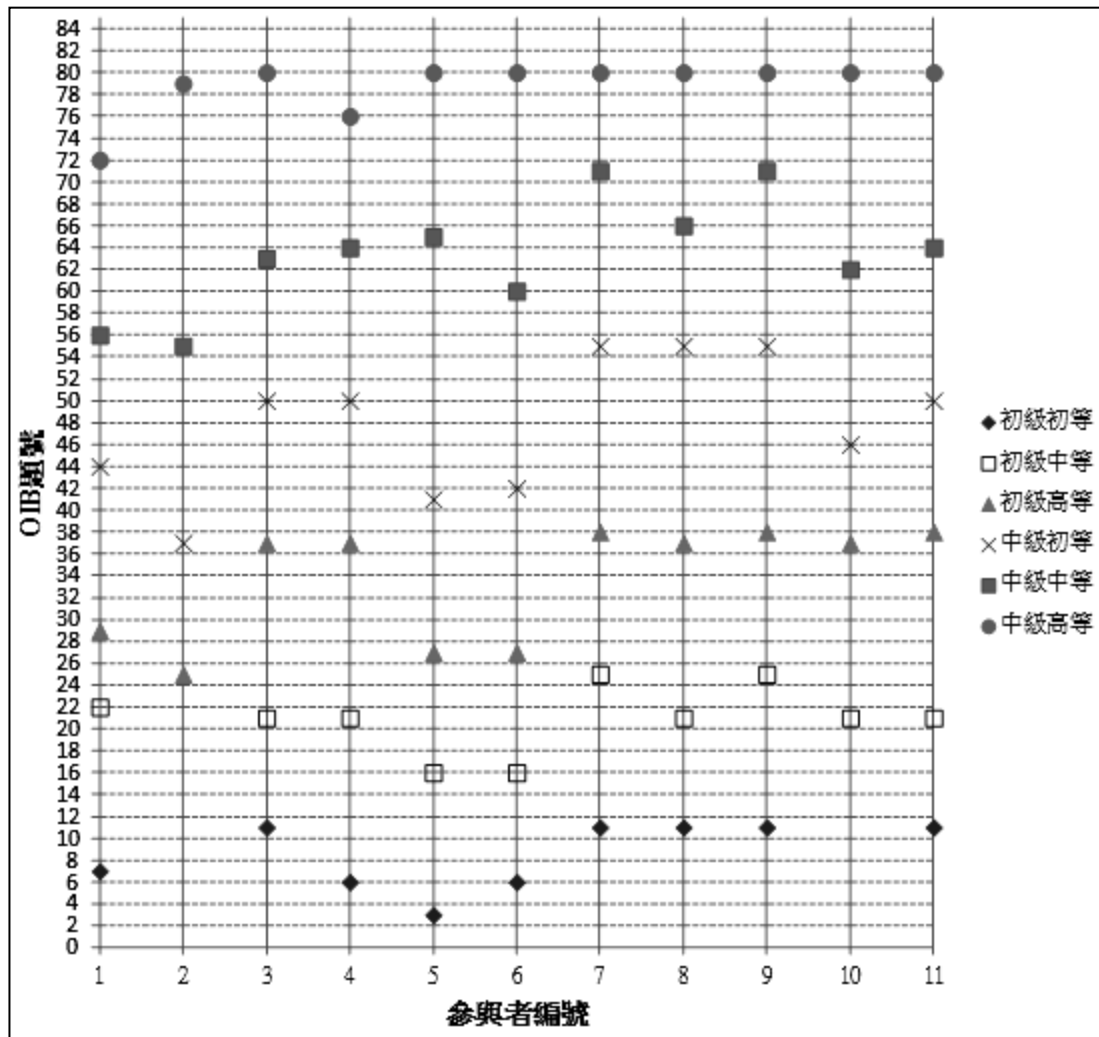


圖 2 第二輪標記結果

4.2 標準設定結果之效度

完成兒童閱讀測驗與 ACTFL 對應的標準設定程序後，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），在此提供程序性效度及內部效度檢核結果。

首先，程序性效度方面，兒童閱讀測驗對應至 ACTFL 的標準設定過程，均依照既定的步驟進行，判定後的討論與會成員之間亦進行充分的討論與分享，如第一回合標記後的討論約進行了兩至三個小時，交換彼此判定的理由與依據。且會議後的問卷調查結果（表 8）各題目的平均數都達到 3.0 以上，同意百分比達 80% 以上；第 1 題至第 3 題平均數甚至高達 3.9，同意百分比高達 98%。此一結果顯示成員了解標準設定流程與書籤法進行方式，且了解 ACTFL 指標各等級能力描述的意涵等研究進行所要求的相關資訊，以上皆顯示標準設定結果具有程序性效度。

表 8 標準設定會議問卷調查結果²

問卷題目	平均數	同意百分比
1. 我了解本次 CCCC 與 ACTFL 指標對應研究會議的目的。	3.9	98%
2. 會議帶領者對於標準設定研究的流程說明得很清楚。	3.9	98%
3. 會議帶領者對於本研究 Bookmark 法的進行方式說明得很清楚。	3.9	98%
4. 會議帶領者對於 ACTFL 的架構和能力描述說明得很清楚。	3.8	95%
5. 我了解 ACTFL 各等級能力描述的意涵。	3.8	95%
6. 研究者所提供依試題難度排序的題本，符合我認為的試題相對難度。	3.3	82%
7. 我是根據 ACTFL 各等級的能力描述，來標記各等級學習者答對率為 67% 的試題編號。	3.8	95%
8. 提供試題對應能力值，有助於我標記各等級學習者答對率為 67% 的試題編號。	3.5	89%
9. 提供試題對應量尺分數，有助於我標記各等級學習者答對率為 67% 的試題編號。	3.3	82%
10. 第一回合後團體中的討論和分享，有助於我進行第二回合的判斷。	3.8	95%
11. 我對於自己所設定的通過門檻有信心。	3.5	89%

內部效度則由每一輪判定結果之全距與四分位差作為依據。六個等級在第二輪的全距均小於第一輪；在四分位差的部分，除了初級初等外，其餘五個等級在第二輪的四分位差均小於第一輪，且在第二輪初級初等到中級高等六個等級四分位差都低於 0.3。這些數據都顯示經過討論後，成員間的判斷趨於一致，可做為標準設定結果的內部效度證據。

4.3 指標對應結果與等級難度對應

經過討論與設定後，本研究的結果如表 9 所示，根據專家們的判定，兒童華語文能力測驗於萌芽級（Pre-A1）的閱讀能力，也就是僅擁有部分 A1 能力的考生能力近似於 ACTFL 指標中的初級初等、初級中等。A1 程度學習者的閱讀能力則相當於初級高等與中級初等兩個次等級。而 CEFR 所描述的 A2 閱讀能力表現，經判定與中級中等和中級高等相近。

² 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 9 CEFR、CCCC 與 ACTFL 等級對照表

兒童閱讀測驗等級	CEFR 等級	ACTFL	
		等級	能力值
茁壯級	A2	中級高等 (Intermediate High)	-1.03
		中級中等 (Intermediate Mid)	-1.75
成長級	A1	中級初等 (Intermediate Low)	-2.42
		初級高等 (Novice High)	-3.14
萌芽級	Pre-A1	初級中等 (Novice Mid)	-3.87
		初級初等 (Novice Low)	-4.91

5. 討論與建議

5.1 各等級典型能力值結果

前面研究結果已提到，各等級典型能力值的四分位差，在第二輪都降至 0.3 以下，兩輪判定四分位差變化最大的等級為中級高等。研究者推測可能是由於參與者對於中級能力描述中「簡短連貫文本」理解資訊程度的定義歧異較大所致，詳見下節。經過第一輪討論後，與會者取得判斷的共識，因此第二輪的四分差降低許多。

根據收集到的口語研究文獻及研究者對於能力描述的分析，CEFR 的 A1 至 B1 約相當於 ACTFL 的初級與中級，也因此採用兒童華語文能力測驗做為媒介，來探討 CEFR 與 ACTFL 初級閱讀能力的對應關係。本研究結果，低於 CEFR A1 以下的 Pre-A1 約與 ACTFL 初級初等、初級中等相當，而 A1 與初級高等、中級初等相當，至於 A2 則與中級中等、中級高等相當。此一結果大致上與先前研究結果及研究者分析能力描述後的預設相近，惟在中級中等和中級高等的判斷結果較高。

根據 Martínez Baztán (2008)、Bärenfänger and Tschirner (2012) 對應口語能力指標的研究結果，前者 CEFR 的 A2 等級與 ACTFL 的初級高等、中級初等、中級中等相當，B1 與中級中等、中級高等相當；後者 CEFR 的 A2 等級與 ACTFL 的中級中等相當，B1 與中級中等、中級高等相當。關於口語能力相關研究與本研究在 A2 能力範圍的對應結果落差，研究者認為除了產出型語言能力與接收型語言能力可能在本質上有所差異外，書籤法於研究流程上的限制也是另一項可能原因。

5.2 各等級之判別與討論焦點

與會專家在進行 OIB 題本等級判定的討論中，都需表述個人的判定歷程與

策略，在採用 ACTFL 能力描述中的各等關鍵特徵，對比試題文本內容及答題關鍵的策略時，幾乎所有的與會專家都表示，他們不以單題而是以鄰近數題的共同特徵為考量；若他們注意到具某等級代表性的特徵開始出現在題目中，便會進一步觀察該特徵是否在相鄰數題間密集地成為答題的關鍵能力，若某區段的試題都要求受測者具備某項關鍵能力方能答對，該區段的前後數題多會成為與會專家定下標記的切點，或成為稍後的討論中受到關注的位置。以下將指出標準設定當天，數個與會專家於討論中針對各等判斷所提出的數個等級轉換的關鍵特徵。

在初級之下三個次等級的討論中，對於詞彙的辨認能力和以對照方式作答的特性，是參與者判斷初級初等與初級中等兩個次等級的關鍵特徵，因此，在有視覺輔助的情況下，依賴單一詞彙的辨認，或直接對照題目與題幹中相同漢字即能做答的題目，在判定過程中被認為是典型初級初等程度學習者的能力範圍。接下來，討論重點從單一詞彙的辨識，提升至大致理解成句文義的能力，對於掌握答題關鍵的能力，也從「對照」提升至簡單的「理解」，凡要求受測者具備以上能力方能做答的題目，都被判定為初級中等至初級高等的能力範圍。

參與者對於初級高等過渡到中級初等的能力判定，以文本的連貫程度為主要討論焦點，與會者認為，具備初級高等典型能力的讀者雖然能閱讀簡單但具完整內容成篇文本，如時刻表或常見的公告，但無法理解連貫的文本，因此這個次等級的閱讀表現仍以在文本中找出與答題關鍵有關的單句，並理解其句義為主，但到了中級初等的典型能力範圍，與會專家的討論便大多集中在連貫文本中需依靠上下文文意以理解部分內容的題目。本次會議需於 OIB 題本中決定的最後兩個切點等級為中級中等與中級高等，在中級階段對學習者要求理解連貫文本的基礎上，這兩項次等級的討論多著墨於情境與文本理解之上，與會者往往分析部分題目的主題是否對兒童受測者而言為最直接、最熟悉的個人生活情境，同時考量答題關鍵是否要求對於文本通篇且全面的理解，或試題對理解文本細節的要求程度；若試題對全面性篇章理解的要求較高，或答題關鍵涉及文本中被視為細節的資訊時，與會者便會傾向做出中級高等的切點判定。

5.3 應用參考

本研究的結果為 CEFR 與 ACTFL 兩個能力指標的初級閱讀能力描述做出客觀的連結，CEFR 能力指標中所定義的「初級使用者」閱讀能力與 ACTFL 自初級中等至中級高等的能力相當。同時，在與會專家進行各等級書籤法設定的討論中，也凸顯出兩個指標在定義初級閱讀能力時，在數項特徵和學習進程上具有共通性。此研究結果可提供初級閱讀教學者之參考。

首先，初級使用者所能掌握的字詞、主題與語言使用情境都以個人熟悉並與生活經驗高度相關者為範疇，文本則需在主題或形式上使內容具可預測性。而學習者在此階段的語言能力進程在兩項語言能力描述中也展現相近的輪廓；初級階段的閱讀能力發展歷程，是由依靠個別詞彙理解文本的階段過渡到理解單一文句，最後再發展出閱讀簡單、連貫文本的能力。

此外，以 CEFR 或 ACTFL 的分級來設計課程內容或語言測驗者，通過本研究結果，可清楚了解 CEFR 與 ACTFL 分級的對應情形。以實施本次書籤法的兒童華語文能力測驗為例，研究結果讓受測者除能以測驗所參照的 CEFR 能力指標分析自身語言能力外，對於參與以 ACTFL 能力指標所規劃之華語課程的受測者與教學者而言，也能利用熟悉的 ACTFL 分級概念解讀測驗結果；在受測者學習歷程不盡相同的現實情況下，仍能客觀地為學習者提供語言學習的進程與實際語言應用能力的資訊，可以實現語言能力測驗不受課程與教材等因素影響的理想。

引用文獻

- 國家華語測驗推動工作委員會 編，2014，《兒童華語文能力測驗學習手冊》，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 編，2015a，《兒童華語文能力測驗技術報告—2013（5）：聽力、閱讀測驗信效度》，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 編，2015b，《華語文能力測驗技術報告—2013（1）：閱讀測驗信效度》，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 編，2016，《兒童華語文能力測驗技術報告—2014（5）：聽力、閱讀測驗信效度》，新北市：國家華語測驗推動工作委員會。
- 莊永山、夏崇舜、尚惠芳 譯，2007，《歐洲共同語文參考架構》高雄市：和遠圖書資訊出版社。
- 陳怡靜、趙家璧，2012，〈寓試於樂—兒童華語文能力測驗〉，《華文世界》109:26-32，台北市：華文世界雜誌社。
- American Council on the Teaching of Foreign Languages. 2012a. ACTFL Proficiency Guidelines. Retrieved on 1 May, 2014, from http://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf
- American Council on the Teaching of Foreign Languages. 2012b. General Preface to the ACTFL Proficiency Guidelines 2012 Retrieved on 1 May, 2014 from <https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012/chinese/simplified-characters/%E9%98%85%E8%AF%BB>
- Angoff, William H. 1971. Scales, norms, and equivalent scores. In Thorndike, R. Ladd. (ed.), *Educational measurement (2nd ed.)* 508-600. Washington, DC: American Council on Education.
- 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.
- Berk, Ronald. A. 1986. A consumer's guide to setting performance standards on criterion referenced tests. *Review of Educational Measurement* 56.1:137-172.

- Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge, UK: Press Syndicate of the University of Cambridge.
- Impara, James C. and Barbara S. Plake. 1997. Standard-setting: An alternative approach. *Journal of Educational Measurement* 34:353-366.
- Kane, Michael. 1994. Validating the performance standards associated with passing scores. *Review of Educational Research* 64:425-461.
- 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, Boulder, CO.
- Martínez Baztán, Alfonso. 2008. *La evaluación oral: una equivalencia entre las guidelines de ACTFL y algunas escalas del MCER*. (doctoral thesis) Universidad de Granada.
- Rasch, Georg. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Institute of Educational Research. (Expanded edition, 1980. Chicago: The University of Chicago Press.)
- Tannenbaum, Richard J. and Caroline E. Wylie. 2008. *Linking English-Language Test Scores onto the Common European Framework of Reference: An Application of Standard-Setting Methodology* (TOEFL iBT Research Rep. No. RR-08-34). Princeton, NJ: Educational Testing Service.

Aligning ACTFL and CEFR proficiency levels: utilizing children's Chinese competency certification

Chia-Pi Chao, Pei-Jiun Lan, Po-Hsi Chen

Steering Committee for the Test Of Proficiency-Huayu

Abstract

ACTFL Proficiency Guidelines (ACTFL) and Common European Framework of Reference for Language (CEFR) are two major language proficiency guidelines adopted internationally in the fields of language teaching and assessment. This study aims to link the framework of CEFR and ACTFL by utilizing the Children's Chinese Competency Certification (CCCC). 11 experts have participated in a two-round standard setting procedure applying the Bookmark Method. The CCCC Reading is a standardized test designed based on the CEFR proficiency descriptions for Basic User (A1-A2). The standard setting procedure was performed on the CCCC Reading focus on ACTFL novice and intermediate levels. The research suggests the following results. First, the pre-A1 reading proficiency which is lower than CEFR A1 is equivalent to the novice low and novice mid levels in ACTFL. The CEFR A1 reading proficiency is equivalent to ACTFL novice high and intermediate low levels and the CEFR A2 is equivalent to ACTFL intermediate mid and intermediate high levels. Second, the evidence of procedural validity and internal validity are provided to support the validation of the standard setting. Finally, in describing reading proficiency and learning progress in the beginner levels, the process of standard setting has revealed similar characteristics in both the CEFR and the ACTFL framework. The results of this study allow the test takers of CCCC to interpret test results with the language proficiency framework which they are familiar with. It can also provide language learners and educators with a more objective correspondence between CEFR and ACTFL.

Keywords: CEFR, ACTFL Proficiency Guidelines, Children's Chinese Competency Certification, standard setting