

華語口語能力電腦化評核工具開發

—以華語文快篩口語測驗系統為例

馬千惠、廖才儀、梁綺容、陳柏熹
國家華語測驗推動工作委員會

摘要

為因應海內外各企業即時評量外籍人才的華語能力，以及學院校等教學單位進行華語能力分班、追蹤外籍生華語學習成效等需求，華測會於 2021 年開始研發，並於 2023 年正式推出「華語文快篩口語測驗」(簡稱快篩口語測驗)。快篩口語測驗作為一套線上口語測驗工具，能提供考生隨到隨考，並在短時間內提供華語口語能力評估結果。口語測驗屬於「表現測驗」，考生的作答內容為開放性語言輸出，成績須交由評分者針對音檔進行人工給分。為維持評分一致性及穩定性，本會研發了一套快篩口語測驗線上評分系統，系統中提供各等級能力指標、評分原則及標準音檔等作為給分參照標準，也設置每半年一次的線上自評檢核機制，以定期檢核評分者的給分穩定性，通過審核者使能繼續參與測驗評分。本文以「華語文快篩口語測驗」為例，闡述華語口語能力電腦化評核工具的研發歷程，首先簡介測驗架構、測驗流程及題型，接著說明本測驗如何在兼顧快速評估口語能力結果的同時，確保一定程度的評分信度，最後介紹測驗成績計算以及結果呈現方式。本研究通過開發華語文快篩口語測驗系統，探索在科技輔助下的評核方式，結果可見本系統讓有快速檢測口語能力需求的教學單位及企業，以更有效獲得口語語言能力評估結果。

關鍵詞：華語能力測驗、口語測驗、電腦化測驗、系統研發

1. 前言

「華語文口語能力測驗」是一套由國家華語測驗推動工作委員會(以下簡稱華測會)負責研發，專為母語非華語者所設計之標準化語言能力測驗。為因應海內外各企業即時評量外籍人才的華語能力，以及學院校等教學單位進行華語能力分班、追蹤外籍生華語學習成效等需求，華測會除了持續推行華語文口語能力測驗正式考試，亦於 2021 年開始研發，於 2023 年正式推出「華語文快篩口語測驗」(以下簡稱快篩口語測驗)。該測驗可供考生隨到隨考，並能在短時間內提供華語口語能力的評估結果。使用快篩口語測驗的單位可隨時為考生開立帳號，考生登入線上測驗系統並完成測驗後，作答音檔將通過使用單位自評或系統自動指派的評分人員進行評分。測驗結果包含量尺分數及評估等級，並呈現與歐洲共同語文參考架構(CEFR)、美國外語教學協會能力指標(ACTFL)，

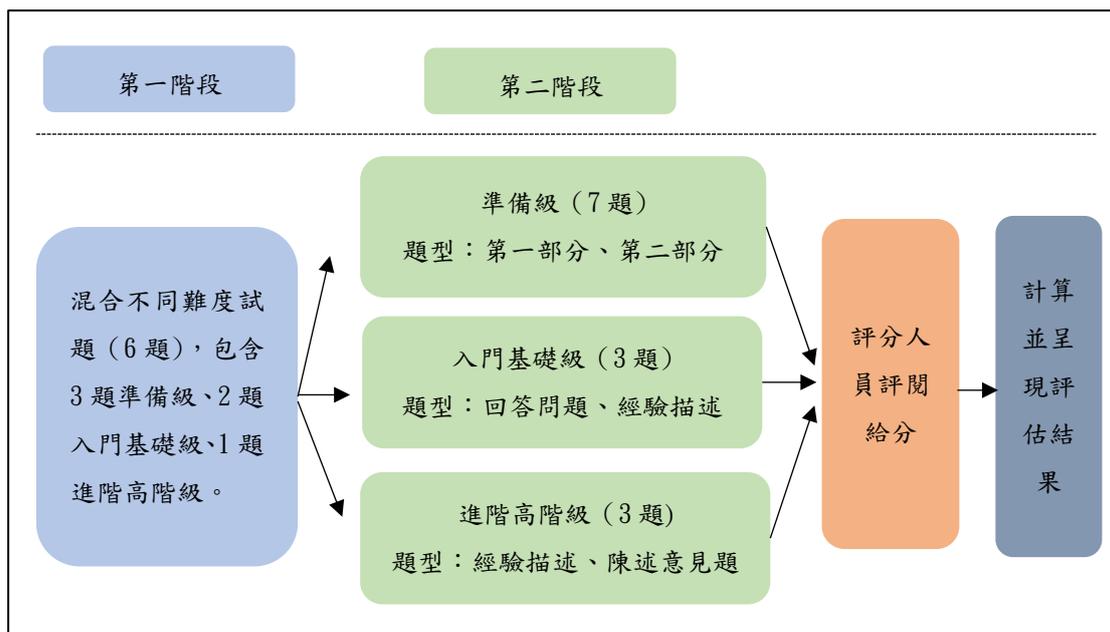
以及臺灣華語文能力基準 (TBCL) 等國際語言能力標準的對應關係。本文首先闡述快篩口語測驗架構、測驗流程及題型，接著說明評分方式、成績計算以及測驗結果。

2. 華語文快篩口語測驗簡介

本節簡介快篩口語測驗的測驗等級與架構，接著依序說明「準備級」、「入門基礎級」、「進階高階級」的題型設計概念。

2.1 測驗等級與架構

華語文口語測驗為一套標準化的電腦化口語測驗，測驗架構主要以 CEFR 為依據，採四等八級架構，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)、精通級 (Level 6) 共八級。以上四等八級已於 2013 至 2018 年間分別完成與 CEFR、ACTFL 兩大語言架構的對應研究 (國家華語測驗推動工作委員會，2013、2014、2017；廖才儀等，2022)，可精準評估語言使用者在使用中文的目標語區域進行日常生活、旅遊、求學、工作等一般口語溝通情境的口語表達能力。建立在華語文口語測驗研發成果的基礎上，參酌 CEFR 於 2017 年公布之新版指標架構，快篩口語測驗提供「準備級」、「入門進階級」與「進階高階級」三個測驗等級，以滿足海內外各企業、教學單位快速檢測華語文口語能力之需求。有別於正式考試的分等測驗架構，快篩口語測驗採用「兩階段測驗架構」，於測驗第一階段考生須作答混合不同難度的試題共 6 題，包含 3 題準備級試題、2 題入門基礎級試題，1 題進階高階級試題；第二階段再依報考等級或快篩聽力測驗結果，提供符合等級程度之相應試題，題數分別為準備級 7 題、入門基礎級 3 題及進階高階級 3 題。考生需在限定的測驗時間內完成全部作答，測驗過程中，考生可在每題作答結束後聆聽自己的作答音檔，若對作答內容不滿意，可重新作答。測驗結束後，快篩測驗系統會透過評分介面，將作答音檔派送給使用單位自行評分，或由華測會培訓的快篩評分人員評閱給分；待評分人員於評分系統中完成音檔給分後，系統會再進一步根據考生兩個階段的整體表現，判斷其口語能力程度並提供測驗結果。圖一為口語快篩測驗架構流程圖。



圖一：快篩口語測驗架構流程圖

2.2 測驗題型

快篩口語測驗題型比照正式考題設計，以「溝通任務」為導向，命題力求內容之普遍性與真實性，取材皆參照 CEFR 對各等級語言學習者及使用者可處理情境範圍之描述，涵蓋個人資料及喜好、學校或職場情境之日常生活、休閒活動、身體與健康、教育、購物等真實生活情境。研發人員根據 CEFR 就各等級語言學習者在表達能力、互動能力及溝通策略等面向的口語能力表現描述，擬定各等級測驗通過者所應具備的基本語言能力表現，並循此方向設計測驗題型，各等級測驗題型請見表一。

表一：快篩口語測驗題型

等級	題型	準備時間	回答時間	試題說明	試題範例
準備級	第一部分	無	10 秒	根據動畫回答問題。每題播放兩次。	詳見華測會官網： https://tocfl.edu.tw/index.php/test/speaking/list/4
	第二部分	10 秒	10 秒	聽到問題後，回答問題。每題念兩次。	
入門基礎級	回答問題	無	30 秒	以完整、簡短的句子回答問題。	詳見華測會官網： https://tocfl.edu.tw/index.php/test/speaking/list/4
	經驗描述	50 秒	1 分鐘	依據題目設定的情境，描述所處環境中熟悉的日常事物或學習經驗。	
進階	經驗描述	50 秒	1 分 40 秒	依據題目設定的情境，描述個人的經驗、	詳見華測會官網： https://tocfl.edu.tw/index

高階級	事件、夢想等。			php/test/speaking/list/4
	圖片描述	50 秒	1 分 40 秒	能表達對抽象主題，例如音樂、圖片的看法並回答問題。
	陳述意見	1 分鐘	2 分鐘	依據題目設定的情境提出意見及看法，使用相關說明及適當例子，解說並支持自己的論點。

準備級程度 (Band Novice) 的語言學習者需具備的基本能力為「語言基本單位的識別能力」。準備級一級 (Novice 1) 學習者在有準備時間的情況下，能針對與個人切身相關的問題，說出少數日常生活中經常使用的高頻詞彙，如生日、年齡、來自哪裡等。準備級二級 (Novice 2) 學習者同樣是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，如多少錢、他在做什麼等。據此，準備級題型針對「詞彙」及「短句」能力規劃了回答簡單問題，問題內容著重與個人高度相關的訊息，目的在測驗考生能否理解問題，並做出相應的回答。

入門基礎級程度 (Band A) 的語言學習者主要發展描述性能力，需具備「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」的基本能力。入門級 (Level 1) 學習者發展「單句層次描述能力」，能使用熟悉的日常用語與詞彙，進行簡單且制式化的回答，在處理任務情境時，必須在說話者緩慢、清晰的簡單敘述前提下，才能對於有即時需求的事物或熟悉的話題進行回應。基礎級 (Level 2) 學習者的描述能力由單句層次延伸為段落層次，進一步展現出稍具穩定度的「段落層次描述能力」，能針對日常生活中與工作、閒暇和例行事務相關的熟悉話題，處理簡單社交往來，也能進行簡單、直接的想法或資訊交換，例如在商店、郵局、銀行等公共場合的簡單交際、詢問一些基本的信息，如問路、指引方向等。據此，入門基礎級的題型架構針對「單句層次描述」和「段落層次描述」的能力分別規劃了回答問題和經驗描述題兩大題型，其中回答問題內容著重與個人生活密切相關的信息，例如工作、閒暇及日常例行生活等，屬於可以直接反應作答的題目，因此不提供準備時間，我們期待考生在聽到題目後以完整的句子作答；經驗描述題的內容則著重個人所處環境中熟悉的日常事物或經驗，此部分提供 50 秒的準備時間讓考生回想自身經驗、組織作答內容。

進階高階級程度 (Band B) 的語言使用者需具備的基本能力為「針對自身經驗或感興趣的話題，做出直接、有次序地描述或提出看法」。進階級 (Level 3) 學習者可以直接且連貫地描述與個人相關的經驗、感覺、夢想、希望等，隨著段落層次組織及連貫能力的發展，這個等級的學習者針對個人感興趣的一或多個不同主題，例如談論音樂、影片等廣泛性的抽象文化主題、發表自己的旅遊計畫、對不同方案進行比較等，可以有次序地說明計畫或事件，並提出簡短的理由支持自己的看法。高階級 (Level 4) 學習者在描述或說明的能力都更加成熟，這個等級的學習者對於一般性議題或有爭議的內容，都能

夠發表清楚、有組織且連貫的個人論點，並且使用適當的例子和相關細節來延伸說明。據此，進階高階級的題型架構針對「連貫性描述」和「陳述個人看法、論點」的能力規劃了描述題和說明題兩大題型。描述題可分成經驗描述題與圖片描述題，目的在測驗考生能否依據題目設定的情境，描述個人的經驗、事件；以及能否客觀地描述圖片上的事件，且與個人經驗做連結對照。說明題又稱為陳述意見題，可分成「選擇並比較不同方案、提出經驗」，以及「對特定議題表達贊成或反對立場」，目的在測驗考生面對不同任務情境時，能否根據試題所提供的資料提出自己的意見和想法，並提出理由支持自己的論點。

綜上所述，快篩口語測驗藉由不同的題型設計，以求在完整的測驗時間內，全面考察考生從簡單的基本生存需求到個人生活、社交互動、求學及工作等不同場景的口語應對能力，通過考生從詞彙到段落層次的整體語言表現，最終評估其在以中文作為目標語的環境對他人有效傳遞需求、交換信息及表達個人觀點的口語能力。

3. 標準化評分

口語測驗屬於「表現測驗」(performance assessment)，考生的作答內容為開放性語言輸出，成績須交由評分者逐一針對音檔進行人工給分。為避免評分者的主觀性影響考生能力判定結果，研發人員考量各等級測驗評量重點、語言能力表現特性、語言任務性質差異等因素，制定出一套標準設定後的評分原則(廖才儀等, 2022)，以確保評分一致性。此外，也制定了標準化評分程序，並透過快篩系統定期檢核評分者的評分穩定度，以確保測驗結果的信度。

3.1 評分原則及標準音檔

快篩口語測驗採用有向度的整體式給分，評分向度包含「內容組織」、「表達能力」及「語言運用」，在評分階段時，評分者需以評分原則中的規範內容為主要給分依據，在綜合考量音檔三個向度的語言表現後給予一個整體級分。其中，「內容組織」考察的是任務完成度、話語組織性和連貫性；「表達能力」考察整體語音被聽者理解的程度，包含回答內容中的語速穩定性、發音可理解性、停頓的頻率及每次停頓的時長等；「語言運用」則考察詞彙及語法結構的豐富性，以及使用適切性等。由於準備級與入門基礎級的溝通任務較單純，音檔內容多半是簡單的詞彙、短語或句子，因此這兩個等級的評分級距為 0 到 3 級分；而進階高階級的作答內容須考量的因素較為複雜，例如話語結構組織性、連貫性等，因此進階高階級的給分級距為 0 到 5 級分。在快篩系統的評分介面中，除了提供評分原則，也提供輔理解評分原則內容的「標準音檔」。這些音檔展示了不同題型在各級分的典型語音表現，使評分者能將評分原則與標準音檔進行對照理解，從而校準評分標準，提升評分的一致性。

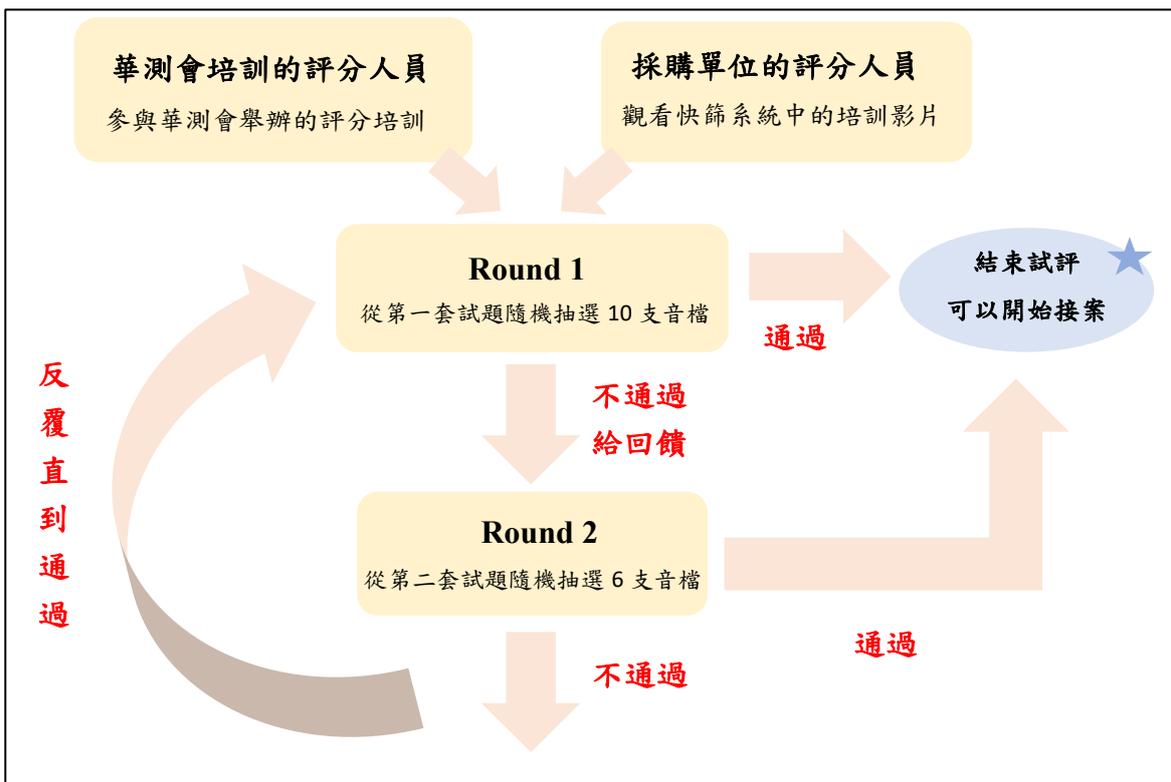
3.2 評分程序

基於華語文快篩測驗快速檢測之目的及需求，快篩口語測驗提供兩種評分人員選派方式供使用單位選擇：

(一)單位自評。使用單位可由單位內部自行挑選評分人員，透過「單位自評系統」進行評閱給分，單位自評系統中提供評分原則、標準音檔等資料作為參照標準，評分時程由使用單位自行決定。

(二)系統自動委派評分人員。由系統自動委派華測會培訓之合格評分人員進行評閱給分，無論實際考生人數多寡，評分時程都可在 7 個工作天內完成，產出測驗結果。

為了兼顧短時間快速評估口語能力的目標，同時維持一定程度的評分信度，無論是使用單位自行評分，還是系統自動委派由華測會訓練的評分人員，均需透過快篩口語系統的評分介面進行音檔評閱。評分介面提供了評分原則、標準音檔以及評分說明短片等系統性評分工具，以輔助評分者達成一定程度的評分一致性。評分人員被告知在音檔評閱過程中，需以評分原則為給分的主要依據。除了上述系統性評分工具，另透過評分介面端設置了每半年一次的線上自評檢核機制，以控管快篩口語評分者的評分穩定性。系統會根據快篩評分者的簽約起始日，每隔半年暫時關閉評分者的評閱權限，並自動提醒快篩評分者進入「自評檢核區」，接著，系統會隨機派發一批音檔供評分者進行線上評核，完成評核後可馬上知道是否通過評核標準。通過評核者可繼續參與快篩口語評分工作；未通過評核者，系統會繼續隨機派發下一批音檔直至快篩評分者通過線上評核為止。圖二為快篩口語評分人員自評檢核流程圖。



圖二：快篩口語評分人員自評檢核流程圖

4. 能力評估結果

評分階段結束後，系統將整合兩階段成績來判斷考生的口語能力水準，評估結果提供「量尺分數」和「快篩評估等級」。由於快篩口語測驗採兩階段測驗架構，考生在兩個

測驗階段中會接觸到跨難度等級的試題，例如報考快篩口語測驗進階高階級的考生，測驗總題數是 9 題，但其中有 5 題會是準備級和入門基礎級的題目(見圖一)。為了將不同難度等級的試題成績建立垂直等化關係，研發人員於 2022 年進行了「華語文口語能力級分等化研究計畫」，該計畫收集了超過兩百位學習者在第一階段共同難度的混等試題，以及在第二階段分流至不同難度試題的作答音檔與音檔給分結果，將準備級、入門基礎級、進階高階級試題的原始分數轉換至同一量尺，最終計算出測驗量尺分數，及其對應之快篩評估等級。快篩測驗的評估結果將語言能力分成 S1 到 S9，共九個等級，每個等級都有相對應的能力描述，以及往下個等級邁進的未來學習策略建議(詳見附錄一)，同時亦提供與 CEFR、ACTFL 及 TBCL 等不同語言能力框架的對應訊息，讓考生能大致了解自身語言能力在不同測驗大綱下的能力對照結果。圖三為快篩口語測驗與 TCTFL、CEFR 之能力等級對照圖。

TOCFL 快篩口語測驗

ACTFL guidelines	TOCFL- Speedy Screening	CEFR
Advanced Mid	S9	B2
Advanced Low	S8	
Intermediate High	S7	B1
Intermediate Mid	S6	
Intermediate Low	S5	A2
Novice High	S4	A1
Novice Mid	S3	
Novice Low	S2	N/A
	S1	

圖三：快篩口語測驗與 TCTFL、CEFR 之能力等級對照圖

5. 結語

快篩口語測驗自 2023 年推出至今，已授權超過 400 個海內外單位，其中包含國內公私立大學系所、語言中心、國際事務辦事處等相關單位，業界諸如旅遊業、科技業、電子業等企業單位，以及位在世界各地如印度、美國、英國、波蘭、日本等多間海外大學及中文學校，使用人數亦有逐年增長的趨勢。對於學習者來說，快篩口語測驗系統能夠

協助其快速了解自身華語文口語能力於各語言架構下的能力落點，並以客觀方式了解自身華語口語能力表現情形及未來學習方向。對於教學及企業單位來說，則能以更省時、省力的方式獲得具有信度的語言能力評估結果，以加速相關作業流程。未來我們也將持續優化服務，以推動華語口語測驗的普及與提升。

參考文獻

Council of Europe. 2018. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (Companion Volume)*. Accessed online, February 16, 2018.

廖才儀、藍珮君、梁綺容、陳柏熹 (2022)。CEFR 與 ACTFL 大綱口語指標對應研究：以「華語文口語測驗」為媒介。華語文教學研究，19(2)，1-32。
<https://www.airitilibrary.com/Article/Detail?DocID=18118429-202206-202207010016-202207010016-1-32>

藍珮君、王從安、彭淑惠、陳柏熹、曾文璇 (2022)。華語文寫作測驗垂直等化研究。2022 年台灣華語文教學年會暨研討會。<https://tocfl.edu.tw/assets/files/ATCSL2022.pdf>

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。2014。《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. 2014. *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。2013。《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. 2013. *Technical Report of TOCFL 2013-3: Reliability and Validity of the Speaking Test*. New Taipei: SC-TOP]

附件一：快篩口語測驗 S1 至 S9 等級能力描述及學習策略

快篩等級	分級能力描述	學習策略
N/A	無	<ul style="list-style-type: none"> 多練習常用詞彙的發音。 讀出具有標音符號的句子，並注意發音的正確性。
S1	<ul style="list-style-type: none"> 能用常見詞彙回答一些和個人資訊有關的簡單問題。 	<ul style="list-style-type: none"> 多練習使用常見詞彙、短語和簡單句子來回答問題。
S2	<ul style="list-style-type: none"> 能用一些常見詞彙或短語回答日常生活中的簡單問題。 	<ul style="list-style-type: none"> 多練習在聽完一段話後，用自己的話說出聽到的內容。 說話的時候要注意詞彙和句子的正確性。
S3 (A1)	<ul style="list-style-type: none"> 能使用句子回答和個人資訊有關的問題。 	<ul style="list-style-type: none"> 多練習簡單的對話，例如：點餐、問地點或方向等。 多練習用完整的句子回答問題。
S4	<ul style="list-style-type: none"> 能簡單的敘述人物、地點和物品。 能使用句子回答日常生活中大部分的問題。例如：住在哪裡、買了什麼、去了哪裡、做了什麼等。 	<ul style="list-style-type: none"> 多練習使用連接詞，連接多個句子，描述自己的經驗。 多練習有組織的說出事件的時間先後和因果順序。
S5 (A2)	<ul style="list-style-type: none"> 能使用句子敘述個人背景、日常生活中熟悉的事物及每日例行性事務。 能簡單地敘述經驗或事件。 	<ul style="list-style-type: none"> 多練習用多種句型，有組織的表達自己熟悉或感興趣的話題，例如：經驗、感覺、夢想等。 針對熟悉的主題，多練習使用數個連續的句子，明確的表達個人立場或觀點，並說明理由。
S6 (B1)	<ul style="list-style-type: none"> 能就一般常見的話題，使用連貫的句子，來敘述經驗、感覺、夢想等。 能有條理的說明計畫或事件；能提出簡短的理由支持自己的看法。 	<ul style="list-style-type: none"> 針對常見的話題，多練習說出話題中的細節，例如：事件發生的原因、發展過程、對自己的影響和感受、為什麼會有這樣的感受等。 針對各種常見需要表達想法的問題，多練習明確的提出自己的看法，並舉例支持自己的論點。
S7 (B1+)	<ul style="list-style-type: none"> 能就一般常見話題，清楚、流暢地敘述經驗或事件，以及當中的細節與解釋個人的感受。 能清楚、詳細地說明計畫及預計採取的相關步驟；能明確地提出看法及論點，並清楚、詳細地說出想法的要點。 	<ul style="list-style-type: none"> 針對生活中的議題，多練習先簡短說明該議題的內容，接著結合自身經驗與觀點，舉出相關例子陳述自己的想法。

快篩等級	分級能力描述	學習策略
S8 (B2)	<ul style="list-style-type: none"> • 能長篇且成段地描述感興趣的一般性社會話題、經驗或事件。 • 針對一般具爭議性的話題，能長篇且成段的表達個人立場、觀點，還能舉出相關的例子延伸並支持個人想法。 	<ul style="list-style-type: none"> • 針對一般議題，練習舉出多種例子、或從不同立場舉出反面例子來支持自己的論點。 • 多練習使用不同的詞彙、句型增加話語的豐富度。
S9 (B2+)	<ul style="list-style-type: none"> • 針對一般具爭議性的話題，能描述經驗、事件的全貌和詳盡的細節。 • 針對一般具爭議性的話題，能使用靈活多變的表達方式、舉出多樣的例子或理由來論證立場，並且能歸納重點，做出結論。 	<ul style="list-style-type: none"> • 多觀察流行口語與正式用語的差別，學習依照不同場合調整說話的方式。 • 練習提高表達的流暢度、詞語使用的精確性與豐富度。 • 針對具多元觀點的話題，練習舉出例子或從不同面向來回應他人的觀點。