

華語文能力測驗技術報告—2014(5)

兒童聽力、閱讀測驗信效度

國家華語測驗推動工作委員會 編著

序

中文近年來已經成為許多國家優先學習的第二外國語言，中文教學（又稱華語教學）也成為教育界新興的重要學科。世界各國的華語學習者越來越多，開始學習華語的年齡也持續下降，充分顯示華語的國際影響力與華語教學的未來發展潛力。

良好的華語教學除了要有優良的師資外，還需要仰賴優質的課程設計、教材及測驗工具。臺灣師範大學於 1956 年開始投入華語教學，這五十年來已經累積相當豐富的經驗，成為全球華語教學重鎮。我們也在教育部委託下，於 2005 年成立國家華語測驗推動工作委員會，並積極發展各式華語能力測驗，希望建立具有公信力的標準化華語測驗，成為全球知名的華語能力評量工具。

為了能有效評量出學習者的不同華語文能力水準，並且與國際的語言能力學習暨評量架構接軌，本會特別研發了不同等級的聽力、口語、閱讀以及寫作「華語能力測驗」，以及針對兒童所研發的「兒童華語文能力測驗」。測驗內容參考了歐洲共同語文參考架構，以及學習者的學習時數與華語能力發展，題材涵蓋食衣住行……等多元面向。為了讓學習華語者及華語相關領域工作者能更明白本會發展之各測驗的內容、測驗實施方式與成績公布之標準化流程，以及本年度正式考試之測驗效能評估，我們特別出版這份技術報告供大眾參考。

國家華語測驗推動工作委員會執行長 陳柏熹 謹識

目錄

一、 前言.....	1
二、 簡介.....	2
(一) 能力描述.....	2
(二) 測驗題型與題數.....	4
(三) 通過門檻.....	6
三、 測驗標準化流程.....	8
(一) 試題收集.....	8
(二) 試題修審.....	9
(三) 預試.....	9
(四) 試題分析.....	10
(五) 題庫輸入.....	11
(六) 組合正式卷.....	11
(七) 檢核正式卷與多媒體檔案.....	12
(八) 成績公布.....	12
四、 測驗評估.....	13
(一) 信度.....	13
1. 內部一致性信度.....	14
2. 測驗訊息量.....	14
(二) 效度.....	17
1. 內容效度.....	18
2. 建構效度.....	21
五、 結論.....	34
六、 文獻.....	35

表目錄

表 1 兒童華語文能力測驗能力描述.....	3
表 2 聽力測驗與閱讀測驗分項能力描述(以萌芽級為例).....	3
表 3 聽力測驗新增題型及測驗目標.....	5
表 4 測驗題型與題數分布.....	6
表 5 測驗通過分數.....	7
表 6 兒童測驗正式考試卷難度分布.....	12
表 7 兒童測驗內部一致性信度摘要表.....	14
表 8 兒童華語文聽力測驗雙向細目表.....	19
表 9 兒童華語文閱讀測驗雙向細目表.....	20
表 10 萌芽級試題難度估計分布.....	21
表 11 萌芽級受測者能力估計分布.....	21
表 12 成長級試題難度估計分布.....	22
表 13 成長級受測者能力估計分布.....	22
表 14 茁壯級試題難度估計分布.....	22
表 15 茁壯級受測者能力估計分布.....	22
表 16 試題適配分布.....	25
表 17 萌芽級測驗不適配試題選項描述性摘要表.....	26
表 18 成長級測驗不適配試題選項描述性摘要表.....	28
表 19 茁壯級測驗不適配試題選項描述性摘要表.....	29
表 20 整體模式適配度指標摘要表.....	33

圖目錄

圖 1 測驗標準化流程	8
圖 2 萌芽級測驗訊息量	16
圖 3 萌芽級測驗測量誤差	16
圖 4 成長級測驗訊息量	16
圖 5 成長級測驗測量誤差	17
圖 6 茁壯級測驗訊息量	17
圖 7 茁壯級測驗測量誤差	17
圖 8 萌芽級測驗受測者與題目分布圖	23
圖 9 成長級測驗受測者與題目分布圖	24
圖 10 茁壯級測驗受測者與題目分布圖	24
圖 11 萌芽級兒童測驗驗證性因素分析	31
圖 12 成長級兒童測驗驗證性因素分析	32
圖 13 茁壯級兒童測驗驗證性因素分析	32

附件目錄

附件 1 兒童測驗測驗各等級正式考試成績單樣張	36
附件 2 萌芽級兒童測驗各題因素負荷量及殘差變異量摘要表	39
附件 3 成長級兒童測驗各題因素負荷量及殘差變異量摘要表	40
附件 4 茁壯級兒童測驗各題因素負荷量及殘差變異量摘要表	41

一、 前言

「兒童華語文能力測驗」(Children's Chinese Competency Certification) (以下簡稱兒童測驗)是一套專為七至十二歲母語非華語之兒童學習者所設計的標準化語言能力測驗，由「國家華語測驗推動工作委員會」(以下稱為本會)專責研發。兒童測驗旨在幫助兒童學習者了解自己的華語程度、激勵他們學習的熱忱，故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，豐富有變化，測驗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。

以下將先簡介 2014 年兒童測驗的基本架構如能力描述、內容及各等級通過門檻，並說明本年度改採量尺分數、增加成績單分項描述之變革，接著描述兒童測驗製卷與成績公布之標準化流程，再說明正式考試信效度的分析結果，最後根據各項分析結果進行討論與建議。

二、 簡介

兒童測驗共分為三個等級，由易至難依序為萌芽級(Sprouting)、成長級(Seedling)與茁壯級(Blossoming)。測驗以「歐洲共同語文參考架構」(Common European Framework of Reference for Languages；簡稱 CEFR)之 A1 級(Breakthrough)和 A2 級(Waystage) (Council of Europe, 2001)的語言能力描述及定義為架構，所評量的語言能力皆落在初級語言使用者(Basic User)的範圍之內。其中，成長級和茁壯級分別根據 CEFR 之 A1 級和 A2 級發展而來。與此同時，為了讓初學華語且年齡較小的兒童能夠順利銜接、熟悉測驗方式並建立自信，兒童測驗基於 A1 級能力之簡單語言任務描述再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)。此節將介紹各等級能力描述、題型設計、測驗題數以及通過門檻。

(一)能力描述

兒童測驗以 CEFR 初級語言使用者的能力描述為架構，發展出萌芽、成長、茁壯三級的能力指標，各等級通過測驗者所具備的聽力與閱讀基本能力如表 1 所示。由於兒童測驗是針對兒童學習者所研發，因此，基本能力描述的制定限定在與兒童生活經驗相關的語言使用範疇之內。

為提供受測者語言學習上更具參考價值的評估結果與建議，本測驗於 2014 年開始於正式考試成績單上提供分項能力的說明與評核。兒童測驗成績單上所提供的分項能力說明，是以 CEFR 初級語言使用者的聽力、閱讀能力描述與分項能力描述為基礎，依照各等級、題型所設定的欲測能力，並根據受測者在各項所測能力下的答題情況，以星等呈現受測者於該項能力的表現。表 2 為萌芽級聽力測驗與閱讀測驗分項能力描述內容，至於萌芽、成長、茁壯三級完整的成績單樣張請見附件 1。

表 1 兒童華語文能力測驗能力描述

測驗等級	聽力理解能力	閱讀理解能力
萌芽級	在話語速度非常緩慢、發音清晰，且停頓或重複的前提下，能理解熟悉且基本的詞語，例如個人資訊、數字、顏色、天氣、喜好、問候語等。	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	在話語速度慢、發音清晰時，能理解與日常生活有關的簡短談話，例如與家人的互動、學校生活、購物、穿著、飲食、交通等等。 在聆聽簡短、簡單、發音清晰的宣佈及說明時，能理解其中的要點，例如自我介紹、電話留言、課堂上的宣佈或天氣預報等。	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	能理解有關居家生活、興趣嗜好、旅遊、休閒活動、同儕間的互動等主題的談話。在聆聽發音清晰的說明或錄製片段時，能掌握主旨及重要資訊，例如：介紹、宣佈、廣播、新聞等。	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。

表 2 聽力測驗與閱讀測驗分項能力描述(以萌芽級為例)

測驗類型	分項能力描述
聽力	<ol style="list-style-type: none"> 1. 能聽懂簡單常見的詞彙。 2. 能聽懂簡單的問題。 3. 能聽懂非常緩慢、發音清楚、簡短、簡單的句子。
閱讀	<ol style="list-style-type: none"> 1. 能理解基本的詞語及簡單的句子。 2. 在有視覺輔助的前提下，藉由掌握基本詞語能大致理解簡單的書寫材料。

(二)測驗題型與題數

由於本測驗的目標受測者為七到十二歲的兒童，此階段兒童的認知能力及語言能力尚在發展中，因此研發題型時，需將兒童受測者的發展特性、語言使用範疇及其生活經驗等一併納入考量。首先，兒童專注時間較短(Mckay, 2006)，在一次測驗中，需藉由多元的作答方式與題型的轉換，提升兒童受測者的注意力，以期能客觀地評估受測者真正的語言能力。其次，兒童生活經驗較為有限，故測驗內容所包含的主題、情境、任務與兒童生活經驗相關，避免因為兒童沒有充足的背景知識而影響其作答。再者，兒童的測驗經驗較少，故須藉由兒童熟悉的測驗形式，以降低作答時因不熟悉測驗方式而產生焦慮，進而影響作答的可能性。此外，本階段的受測者為初級語言使用者，整體能力發展相當初階，理解文本的同時往往需要視覺輔助，因此無論是聽力測驗或是閱讀測驗都採用了多種圖文相輔的題型。

本會兒童測驗研發人員(以下簡稱研發人員)綜合考量兒童發展特性、語言使用範疇及生活經驗等因素，依據三等級學習者不同的語言能力描述(見表 1)做為題型設計的基礎，期望透過測驗結果評量受測者是否達到 CEFR 描述中初級使用者的語言能力。在評估過往試題內容與 CEFR 聽力與閱讀之分項能力描述後，兒童聽力測驗與閱讀測驗在 2014 年度進行題型微調，聽力測驗與閱讀測驗各停用一個題型，聽力測驗新增兩個題型，使兒童測驗更符合 CEFR 以語言任務達成與語言溝通能力為核心的精神。

聽力測驗方面停用「完成對話」題型，原因為當考生回答此類題型時，除須理解對話者的本意外，尚須選擇合宜的選項回應，在評量受測者聽力能力的同時還有要求口說能力之虞，故決定刪除之；此外，由於本測驗奠基於 CEFR 能力描述進行測驗設計，為在測驗中更全面地檢視 CEFR 關於 A1 等級使用者「聽宣布及說明指導」之能力描述，決議在原「聽力選圖」的題型中，加入針對該能力設計之試題，並增加 A1 程度之「段落理解」題型以評測此項能力，詳見表 3。

表 3 聽力測驗新增題型及測驗目標

聽力題型	CEFR 能力	測驗目標
聽力選圖	A1 聽宣布 及說明指導	能理解並跟上詳細及緩慢的說明。
段落理解	A1 聽宣布 及說明指導	對於簡短、簡單的宣布或說明，能聽懂其中的要點，例如：自我介紹、電話留言、口頭通知、學校或課堂上的宣佈、天氣預報、注意事項或指令等。

閱讀測驗方面，「選詞填空」題型以考察受測者詞彙語法能力為主，而兒童測驗的設計以評量初級使用者語言能力為重點，詞彙與文本內容的選擇範圍有一定的限制，選詞填空的題型在這樣的限制下也易出現考點重覆的問題。因此，為使兒童測驗的所測能力與 CEFR 的精神更貼近，自 2014 年起停用此題型。同時，三等測驗進行等化程序，將試題難度量尺整合後(詳見下節)，「閱讀材料」題型中的部分試題難度符合萌芽級的組卷範圍，經過試題內容與提問方式確認與萌芽級閱讀測驗欲測能力相符後，亦決議將閱讀材料題型納入萌芽級的組卷範疇。

微調後的兒童測驗，聽力部分有聽力選圖、看圖回答、聽力連連看、會話理解以及段落理解五種題型，閱讀測驗則有看圖辨義、圖文連連看、閱讀材料與短文理解四種題型來測量兒童華語文能力。表 4 為兒童測驗三等級的題型與題數分布，在同一等級內的聽力、閱讀題數皆相同；三等級施測的總題數分別為萌芽級 40 題、成長級 50 題以及茁壯級 50 題，施測時間則分別約為 40 分鐘、50 分鐘以及 60 分鐘。

表 4 測驗題型與題數分布

測驗等級	測驗內容	題型(作答方式)	題數	總題數
萌芽級	聽力理解	聽力選圖 (選擇)	7	20
		看圖回答 (選擇)	8	
		聽力連連看 (配合)	5	
	閱讀理解	看圖辨義 (是非)	12	20
		圖文連連看 (配合)	4	
		閱讀材料 (選擇)	4	
成長級	聽力理解	聽力選圖 (選擇)	7	25
		看圖回答 (選擇)	6	
		聽力連連看 (配合)	7	
		段落理解 (選擇)	5	
	閱讀理解	圖文連連看 (配合)	8	25
		閱讀材料 (選擇)	8	
茁壯級	聽力理解	短文理解 (是非)	9	25
		聽力選圖 (選擇)	7	
		會話理解 (選擇)	8	
	閱讀理解	段落理解 (選擇)	10	25
		閱讀材料 (選擇)	12	
		短文理解 (是非)	13	

(三)通過門檻

兒童測驗三等級中，成長級與茁壯級分別與華語文能力測驗(Test of Chinese as a Foreign Language，簡稱 TOCFL)的入門級、基礎級門檻相當，皆依據 CEFR 之 A1 級與 A2 級發展而來；主要差異為測驗目標對象不同，前者針對兒童設計，而後者則是成人。

過去兒童測驗三等級採各等級內進行水平等化，為了讓不同等級間的試題難度可相互比較，並進一步將試題難度參數連結至 TOCFL 量尺，與同為 A 等的 TOCFL 入門級、基礎級試題難度進行比較，以讓兒童測驗成長級與茁壯級可採用 TOCFL 入門基礎級測驗透過標準設定流程所制定之 A1、A2 通過門檻，本會於 2013 年進行兒童測驗等化程序。

首先，依據新版三等合併預試卷測驗題型與題數分布，從萌芽級、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 1244 人次和閱讀測驗 1231 人次。原先各自獨立的萌芽級、成長級與茁壯級測

驗題庫，將可藉由此一平等化卷試題難度估計結果，再估計出題庫內其他試題難度參數，各等級的試題難度參數可連結至同一把量尺之上，使不同等級間的試題難度得以相互比較。

此外，為了使兒童測驗難度參數連結至 TOCFL 難度參數量尺，另外邀請一群受測者同時作答此一兒童測驗平等化卷與 TOCFL 入門基礎級測驗平等化卷。TOCFL 平等化卷之題力、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也就連結至 TOCFL 難度參數量尺之上。自此，同為 CEFR A 等測驗之兒童測驗與入門基礎級測驗間的試題難度，也有了相互比較的基準。

研發人員依據上述測驗平等化結果及 TOCFL 入門基礎級標準設定結果(國家華語測驗推動工作委員會，2015a；國家華語測驗推動工作委員會，2015b)，制定兒童測驗通過門檻。其中，成長級與茁壯級採用 TOCFL 入門級與基礎級通過門檻，萌芽級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共同訂出。

同時，自 2014 年起，改以量尺分數呈現受測者測驗結果與等級通過判定之依據，量尺分數是將考生能力值透過制定的平均數與標準差進行轉換而得到的測驗分數，此舉能更有效地確保測驗通過標準，不受各次組卷試題間的些微難度變化影響。兒童測驗量尺分數由測驗專家基於過去參與本會所舉辦兒童聽力、閱讀測驗之受測者能力值分布情形，並考量制定量尺分數所需考量的要素(Dorans, 2000)所訂定。三等聽力測驗量尺分數平均數、標準差皆為 75 及 12.5；閱讀測驗則皆為 70 及 10，兩項測驗量尺分數範圍皆為 0-100 分，聽讀合計之量尺分數滿分均為 200 分，萌芽、成長與茁壯級測驗通過分數如表 5 所示。

表 5 測驗通過分數

測驗等級	通過分數
萌芽級	117
成長級	116
茁壯級	118

三、 測驗標準化流程

2014 年度兒童測驗標準化流程主要可分為八個步驟(如圖 1 所示)，依據包括試題的收集、修審、預試、分析、題庫輸入、組合正式卷、檢核正式卷與多媒體檔案，以及成績公布。各部分詳細說明如下所述：

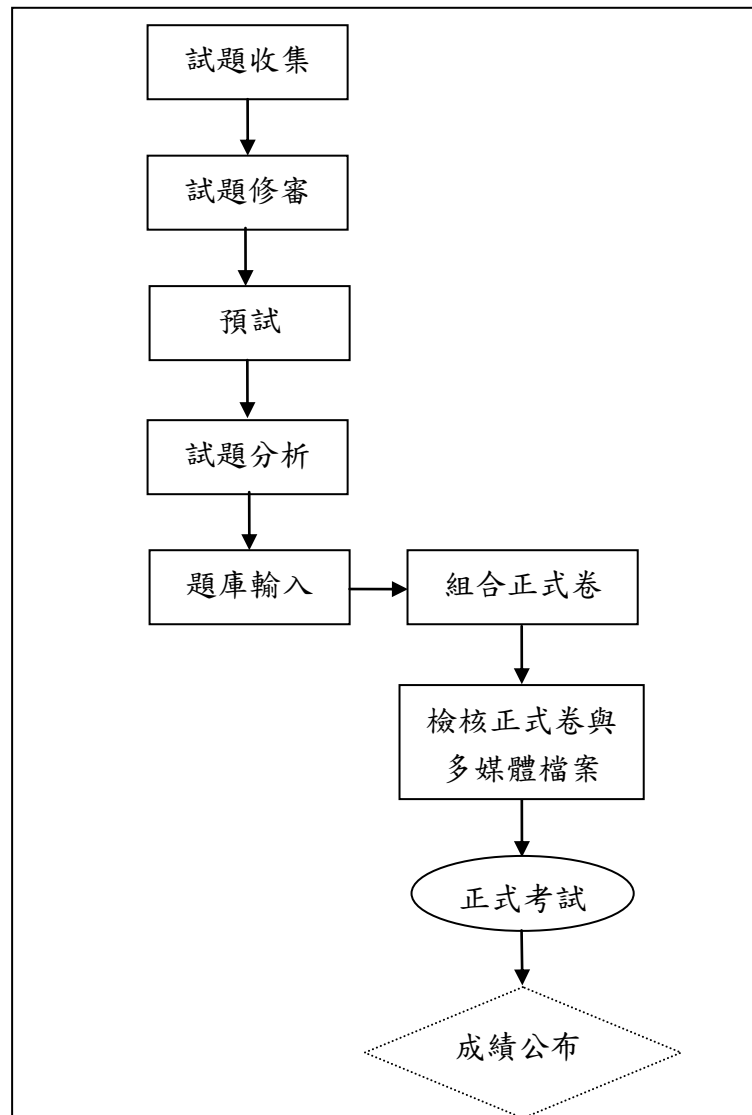


圖 1 測驗標準化流程

(一) 試題收集

2014 年兒童測驗試題收集分為兩階段進行，先舉辦命題研習，並於研習後回收試題進行審查。此次聽力及閱讀測驗共有 18 位老師通過審查，爾後召開的進階命題研討會於會前先由研發人員彙整命題常見問題，再回饋審題意見並加強與命題教師之溝通，進行二次收題。本年度聽力測驗共回收 161 道題目，閱讀測

驗則回收 287 道題目。

(二) 試題修審

1. 會內初審

將試題收集階段完成修訂的試題，交由非兒童測驗研發人員共三人進行第一階段會內初審，並回收審題意見，審題時間為兩週。接著由研發人員根據審題意見修改試題，修改時間為兩週。

2. 專家學者外審

會內初審修改後的試題，在第二階段將交由華語文教學、測驗領域的專家學者及資深華語教師審查並提供意見，此階段審題時間約二至四週。研發人員再依據審題意見修改試題，修改時間為兩週。

3. 製作試題相關多媒體檔案

在進行預試前，審訂完成的試題需製作相關媒體檔案，包含製作測驗說明音檔及試題音檔，繪製並確認試題圖片及題本排版。在測驗說明與試題音檔製作方面，延請專業錄音員進行錄音，並由錄音室製作音檔，過程中研發人員負責確認錄音內容正確性與音檔品質。

此外，考量兒童測驗是為初級華語學習者而設計，為輔助受測者了解考試進行方式，測驗說明音檔的製作採華語搭配其他語言的形式，多國語言測驗說明版本包含中英版、中西版、中法版、中德版、中越版、中韓版以及中日版。

在此最終定稿階段，試題圖片如需更動修改，製作時間將延長一至兩週，待圖片定稿後，再交由美編人員排版，並由研發人員確認內容直至無誤，最後交由印刷廠商印刷預試題本。

(三) 預試

延續 2013 年做法，採三等題型合併之施測形式收集樣本。一份預試卷中包含為萌芽級、成長級及茁壯級受測者設計的試題，並將聽力測驗與閱讀測驗的題本各自獨立，考生可選擇只參加一項測驗或兩項測驗皆參加。這項改變能使每道試題受測樣本的能力分布更為廣泛，降低試題難度參數¹之估計誤差，提高估計

¹ 難度參數指的是由 Winsteps 軟體所估計出的試題難度參數估計值。

精準度；亦可有效避免單一等級試題因預試樣本較少而無法完成收樣的情形。

本年度兒童測驗透過臺灣地區及海外地區收集預試樣本，由於預試樣本來自世界各地，無法進行分層隨機抽樣，取樣方式為便利取樣。收樣時為使樣本具有代表性，盡量控制每一份預試卷的樣本涵蓋亞洲、美洲、歐洲地區考生。此外，亦避免同一預試卷樣本只來自單一地區。總計聽力測驗預試受測者為 760 人次，閱讀測驗則為 642 人次。

(四) 試題分析

將預試階段受測者之作答反應交由統計分析人員進行試題分析，並以試題反應理論(Item Response Theory；簡稱 IRT)作為分析取向。由於兒童測驗中各試題作答反應皆為非對即錯，屬於二元計分方式(dichotomous items)，故採取 IRT 中的 Rasch 模式(Rasch, 1960)進行資料分析，分析軟體為 Winsteps 3.68.2 版(Linacre, 2009)。Rasch 模式如公式(1)所示，

$$P(X_{ni}=1|\theta_n,\delta_i)=\frac{\exp(\theta_n-\delta_i)}{1+\exp(\theta_n-\delta_i)} \quad (1)$$

其中， θ_n 表示受測者 n 的能力； δ_i 表示試題 i 的難度(difficulty)； $P(X_{ni}=1|\theta_n,\delta_i)$ 表示受測者 n 在試題 i 答對的機率。由公式(1)可發現，Rasch 模式的假設為，受測者答對每一道試題機率受到受測者具備的能力及試題難度的影響，將受測者在各試題上的反應，透過公式(1)即可估計出受測者的能力以及試題的難易度，並且讓施測不同試題的受測者其能力可以互相比較。因為，在 IRT 中，受測者的能力參數與試題參數是同時納入在一個模式裡，因此，估計受測者能力時，已經考量了試題參數的影響，估計試題參數時，也考量了受測者能力的影響。

評估預試試題品質的依據為 Winsteps 3.68.2 版輸出報表中的統計指標——訊息加權適配度統計量(inlier-pattern-sensitive fit statistic)之均方(mean-square)，及其標準化 Z 值(z-standardized)，分別簡稱為 Infit MNSQ 及 Infit ZSTD。評估標準為試題之 Infit MNSQ 數值介於 0.7 至 1.3 者以及 Infit ZSTD 介於 -3.0 至 3.0 者，表示試題適配，意即試題品質與測驗研發目標一致、試題品質良好。同時，尚提供試題鑑別度指標²、難度 P 值³及選項分析等資料作為輔助，以更全面的觀點評估

² 試題鑑別度指標為 D 值及點二系列相關係數。

³ 難度 P 值為高低分組受測者答對率之平均數。

試題品質。

(五)題庫輸入

預試後的試題在經由試題分析後，由研發人員參考各評估指標、試題相關資料以及內容等進行討論，以決定試題是否保留。扣除共同試題(或稱定錨試題)後，本年度兒童測驗題庫新進試題題數為聽力測驗 95 題、閱讀測驗 95 題。

(六)組合正式卷

兒童測驗正式卷皆由經過預試且符合試題分析標準之試題組成，並依照固定比例分配各難度區間之題數，自題庫中挑選難度參數符合試卷難度區間要求的試題。組卷題數分布基本原則為切截點附近配置 48% 試題，以確保通過門檻能力值的估計精準度；低於切截點的試題占 40%，降低程度較低考生作答的挫折感；高於切截點的試題占 12%，以區辨程度較高考生。

各測驗等級為因應各等級的通過門檻不同，而在組卷時有不同的主要難度區間範圍，萌芽級聽力測驗與閱讀測驗的難度區間範圍約介於-6.00 至-3.00 logits 之間；成長級聽力測驗與閱讀測驗的難度區間範圍約介於-5.00 至-1.50 logits 之間；而茁壯級聽力測驗與閱讀測驗的難度區間範圍約介於-3.00 至 0 logits 之間。

正式卷初稿組卷完成後，由研發人員檢核各卷試題之主題分布是否過度集中、是否出現重複的考點、是否符合雙向細目表(參見頁 19-20，表 8 及表 9)之規劃等，再根據檢核建議更換試題，直至整份試卷皆無上述問題後始定稿。2014 年度各測驗等級正式考試卷之難度分布如表 6 所示。

表 6 兒童測驗正式考試卷難度分布

試題難度	等級		萌芽級		成長級		茁壯級	
	聽力	閱讀	聽力	閱讀	聽力	閱讀	聽力	閱讀
<-4.99	35%	20%	-	-	-	-	-	-
-4.99~-4.50	25%	45%	4%	4%	-	-	-	-
-4.49~-4.00	30%	25%	24%	12%	-	-	-	-
-3.99~-3.50	-	10%	12%	20%	-	-	-	-
-3.49~-3.00	10%	-	12%	32%	-	-	-	-
-2.99~-2.50	-	-	24%	12%	-	-	-	32%
-2.49~-2.00	-	-	16%	12%	28%	24%	-	-
-1.99~-1.50	-	-	8%	8%	12%	32%	-	-
-1.49~-1.00	-	-	-	-	36%	-	-	-
-0.99~-0.50	-	-	-	-	12%	8%	-	-
-0.49~0.00	-	-	-	-	12%	4%	-	-
>0.00	-	-	-	-	-	-	-	-

(七)檢核正式卷與多媒體檔案

由於兒童測驗採紙筆作答的施測形式，因此，正式卷組卷完成後，必須印刷題本。交付印刷前，研發人員針對題本排版內容進行校對工作，首先為正體版題本的校對，包含確認錄音稿與題本的內容是否一致、檢查圖片與文字的搭配是否正確、漢語拼音與注音符號的標記是否正確、題號與頁數的編碼是否正確等。確認無誤後，再校對簡體版題本，內容及程序同正體版題本。此階段需反覆校對正、簡兩個版本與錄音稿件間的一致性，直至正確無誤。確認題本排版後，即根據錄音稿件完成錄音工作，並與排版完成之正、簡體版題本互相校對，確認錄音品質與內容正確性，確認後試卷檢核即完成，並交付印刷，作為正式考試之用。

(八)成績公布

兒童測驗成績公布流程如下：首先，檢核受測者作答資料並確認無誤後，接著檢核成績報表，確認無誤後即為受測者考試成績，最後進行成績單及證書之印製與寄發工作。

四、測驗評估

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，通常可透過該測驗的信度與效度分析來進行整體性的評估。因此，本節將討論 2014 年度兒童測驗之信效度來說明兒童測驗之效能。2014 年度兒童測驗採聽力和閱讀測驗合併施測的形式，雖然聽力理解能力與閱讀理解能力皆屬於語言能力中的接收能力(receptive skills)，兩者間具有一定程度的相關性，但仍各有其獨特之處，故以下分析將分別描述聽力測驗與閱讀測驗結果。2014 年度正式考試三個等級到考總人數分別為萌芽級 844 位、成長級 852 位，以及茁壯級 593 位。

(一)信度

信度指的是測驗結果的穩定性與一致性，若一份測驗無論在任何時間、任何地點、由任何人進行施測、由任何人計分時，都能得到一致性很高的測驗結果，就表示此份測驗具有很高的信度，也就表示該測驗所獲得的測驗結果(或成績)之測量誤差很小(或精準性高)。以下將從兩大測驗取向—古典測驗理論(classical test theory；簡稱 CTT)及 IRT—分別說明 2014 年「兒童華語文能力測驗」信度。

古典測驗理論是以測驗的信度來評估測量精準度，其假設為接受相同測驗的受測者具有相同測量精準度，因此，測量精準度是以測驗為單位來表示，即施測同一份測驗的受測者其測驗精準度皆相同。CTT 常用來評估測驗信度的指標主要分為四類，分別為：

1. 再測信度，表示在不同時間點施測所獲得的測驗成績是否具有一致性；
2. 複本信度，表示經由施測不同題本所獲得之測驗成績是否具有穩定性；
3. 內部一致性信度，表示測驗所測量到的潛在特質是否具有一致性；
4. 評分者信度，表示經由不同評分者所得到的評分結果是否具有一致性。

而 IRT 的假設則為同一份測驗中的試題，對於不同能力的受測者具有不同測量精準度；當施測的試題難度越符合受測者能力時，對於獲知受測者能力越有幫助，測量精準度越高，反之，當施測試題難度與受測者能力差距越大時，對於了解受測者能力的幫助並不大，測量精準度較低。在 IRT 中，是以測驗訊息量(test information： $TI(\theta)$)的觀點來表示測量精準度，訊息量越高，表示施測試題對於受測者能力的測量精準度越高。

1. 內部一致性信度

兒童測驗為一次性測驗，亦即測驗僅施測一次，所有受測者於同一時間接受同一份測驗，因此，可藉由內部一致性信度來表示測驗信度。當每道試題測量結果的相關性高時，則顯示測驗試題皆測量到相同的潛在特質，即內部一致性高；反之，當測量結果的相關性低，則表示測驗試題測量到其他潛在特質，即內部一致性低。

常用於表示內部一致性的指標有折半信度、Cronbach's α 係數以及庫李 (Kuder-Richardson) 20 號或 21 號公式。此節採用 Winsteps 3.68.2 版所輸出之庫李 20 號公式之係數作為內部一致性信度指標，描述 2014 年度兒童測驗正式考試信度。

表 7 顯示萌芽級、成長級與茁壯級聽力測驗信度係數介於 .86 至 .87 之間，閱讀理解測驗三個等級信度係數介於 .78 至 .84 之間，此數據顯示 2014 年兒童華語文聽力與閱讀理解正式考試卷信度大致良好。

表 7 兒童測驗內部一致性信度摘要表

	萌芽級	成長級	茁壯級
聽力測驗	.86	.87	.87
閱讀測驗	.81	.78	.84

2. 測驗訊息量

在 IRT 中，訊息量的計算公式為：

$$I_i(\theta) = \frac{P_i'(\theta)^2}{P_i(\theta)Q_i(\theta)} \quad (2)$$

其中， I_i 表示試題 i 對於能力為 θ 的受測者所提供的訊息量； $P_i'(\theta)$ 為 $P_i(\theta)$ 對 θ 的一階微分； $P_i(\theta)$ 如公式(1)所示；而 $Q_i(\theta)$ 則為 $1 - P_i(\theta)$ 。將一份測驗中所有施測試題之訊息量加總，即為測驗訊息量。測驗訊息量亦表示為測量誤差 (standard error；簡稱 SE)，如公式(3)所示：

$$SE(\theta) = \frac{1}{\sqrt{TI(\theta)}} \quad (3)$$

其中， $TI(\theta)$ 為整份測驗對於能力值為 θ 的受測者所提供的測驗訊息量。由公式(3)可知，當測驗訊息量越大，對於測量受測者能力的誤差越小，測量精準度

越高。當受測者能力分布近似於常態分布(normal distribution)時，經由 CTT 中測量標準誤與信度的關係式以及公式(3)所獲得的 SE，可換算成相對於 CTT 中的測驗信度，如公式(4)所示：

$$SEM = \sigma\sqrt{1-r} \quad (4)$$

其中，SEM 表示測量標準誤(standard error of measurement；簡稱 SEM)，即為公式(3)所獲得之 SE； σ 表示受測群體能力的標準差； r 表示測驗信度。

圖 2、圖 4 及圖 6 為萌芽級、成長級和茁壯級聽力與閱讀測驗之測驗訊息量，而圖 3、圖 5 及圖 7 則為三等聽力與閱讀測驗之測量誤差。由圖 2 及圖 3 可知，萌芽級聽力測驗正式考試的測驗訊息量在量尺分數為 55 至 58 分時最大，數值皆為 4.49，測量誤差同樣為 5.90，而在萌芽級門檻的測驗訊息量則為 4.43，測量誤差 5.94，信度係數為.90；萌芽級閱讀測驗正式考試之測驗訊息量以量尺分數 54 分為最大，數值為 4.70，測量誤差為 4.62，在通過門檻的測驗訊息量則為 4.60，測量誤差為 4.66，信度係數為.90。上述結果顯示，萌芽級聽力與閱讀測驗在通過門檻所提供的測量相較於其它分數而言，具有較高的訊息量與較低的測量誤差，且在通過門檻具有良好信度。

同樣的現象在成長級聽力與閱讀測驗與茁壯級聽力與閱讀測驗(如圖 4 至圖 7)也可發現，兩等級測驗對於提供門檻之量尺分數的測驗訊息量都較高，測量誤差都較低，且信度係數介於.87 至.92 之間，具有良好信度。如成長級聽力測驗與閱讀測驗通過門檻的測驗訊息量分別為 5.36 與 5.35，測量誤差分別為 5.40 與 4.33，信度係數分別為.92 與.87。而茁壯級聽力測驗與閱讀測驗通過門檻的測驗訊息量則分別為 5.48 與 5.29，測量誤差分別為 5.34 與 4.35，信度係數分別為.92 及.89。

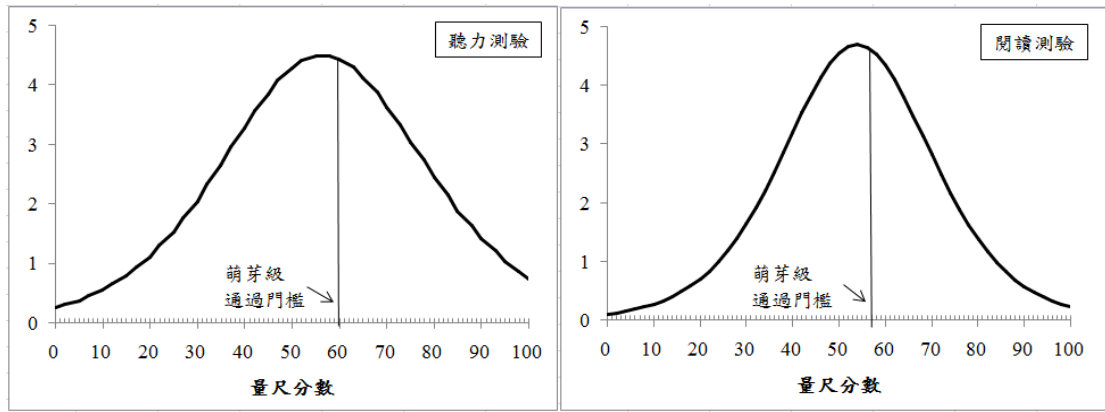


圖 2 萌芽級測驗訊息量

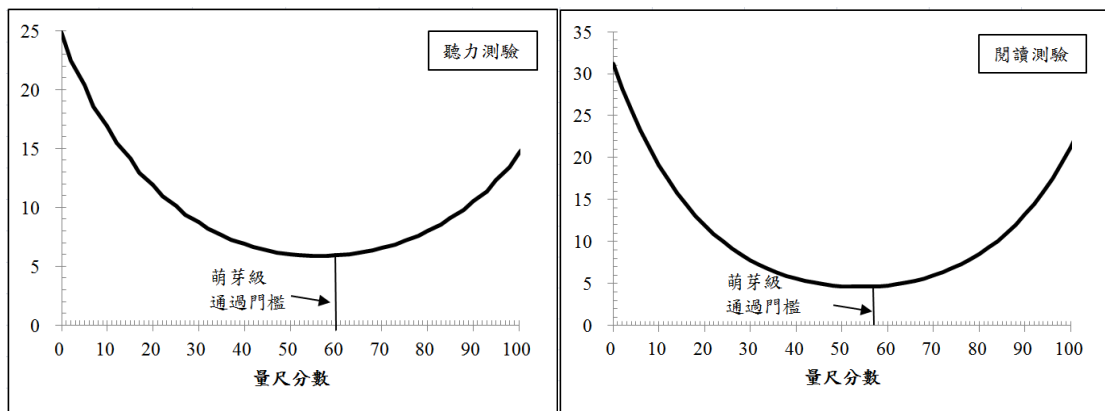


圖 3 萌芽級測驗測量誤差

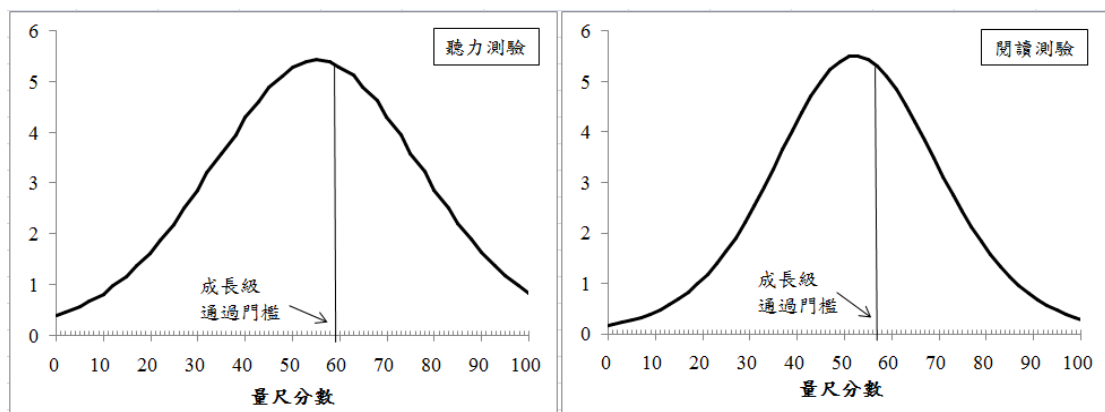


圖 4 成長級測驗訊息量

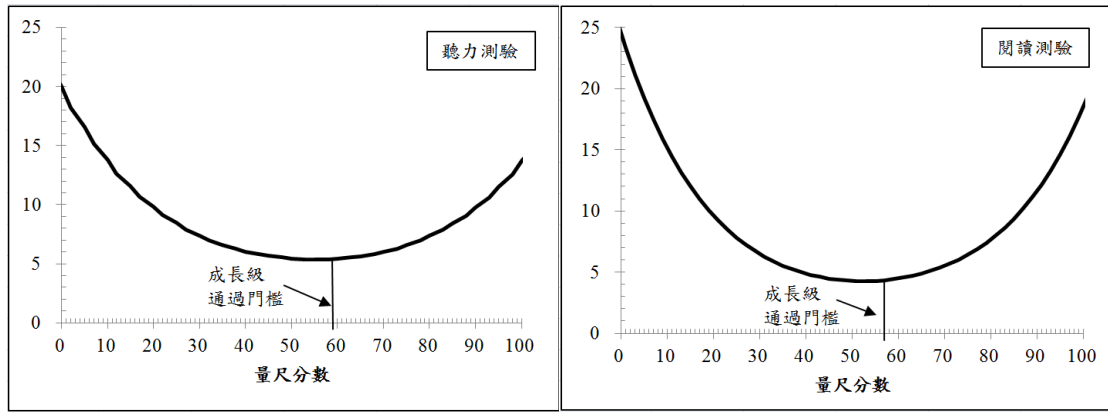


圖 5 成長級測驗測量誤差

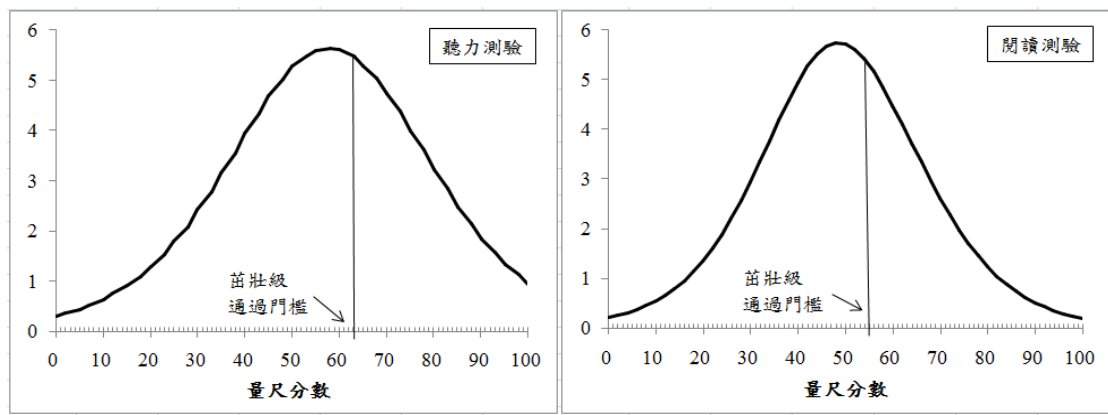


圖 6 茁壯級測驗訊息量

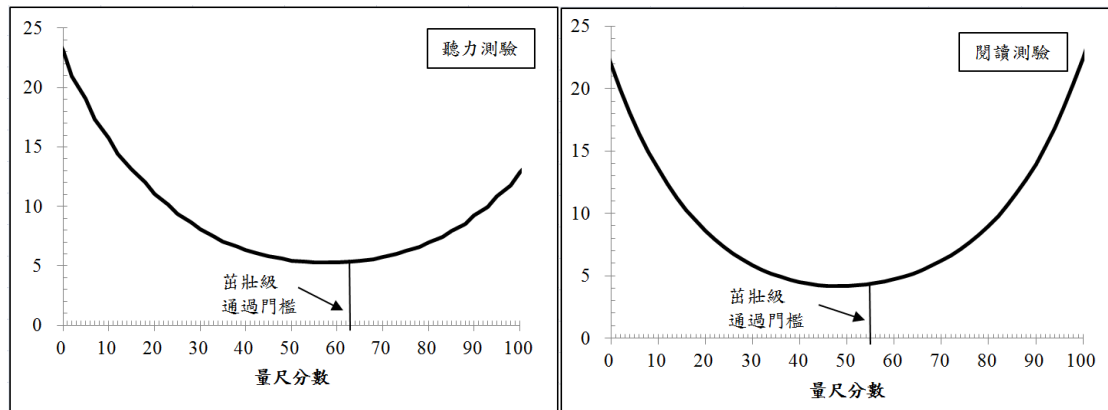


圖 7 茁壯級測驗測量誤差

(二)效度

所謂測驗效度指的是檢驗一項測驗是否能測量到欲測量的能力(或潛在特質)。由於無法直接觀察目標測量能力，因此，測驗效度皆須藉由受測者在試題上的作答反應或行為間接推估。通常用來驗證測驗效度的證據主要分為三大類：

第一，內容效度(content validity)，指的是測驗內容的相關證據；第二，建構效度(construct validity)，即關於測驗架構的證據；第三，效標效度(criterion validity)，指測驗結果預測力的相關證據。內容效度的相關證據主要為評估、分析測驗所測量的能力以及各內容題數比重分配，是否符合測驗所欲測量的能力定義，通常會請所欲測量能力之學科專家評估試題是否符合上述要求，經由專家提供試題內容效度的方式則稱為專家效度；建構效度相關證據指的是評估測驗試題所測量到的能力是否與測驗發展時所定義的架構或內容相同，試題分析以及因素分析常用來提供建構效度的證據；受測者進行測驗的同時，調查受測者相關能力或表現，分析其與測驗結果的相關程度，屬於同時效度，可做為效標效度的一種證據。

以下將從專家效度、試題分析與驗證性因素分析(confirmatory factor analysis)等三方面來描述 2014 年度兒童測驗之內容效度以及建構效度。

1. 內容效度

本會發展兒童測驗之初，集結華語教學、語言學、兒童心理與教育學相關領域專家共同擬訂各等級測驗目標，接著依據各等級目標訂定雙向細目表(如表 8 至 9)，命題教師即根據雙向細目表及相關文件命題。兒童測驗研發人員回收各等級試題後，經測驗標準化流程(如圖 1)中試題修審步驟之第二階段：專家學者外審，聘請語言教學專家與華語教師審查各等級試題是否符合雙向細目表設定的內容，若符合，則表示測驗試題大致測量到測驗目標，此專家學者外審階段提供了兒童華語文能力測驗專家效度。

表 8 兒童華語文聽力測驗雙向細目表

題型 測驗等級	聽力連連看		看圖回答		聽力選圖			對話理解	聽力材料	
	萌芽	成長	萌芽	成長	萌芽	成長	茁壯	茁壯	成長	茁壯
能力描述										
能聽懂非常緩慢、發音清楚、簡短、簡單的句子。	5				7					
能聽懂非常緩慢、發音清楚、簡短、簡單的對話。		7				7				
能因幾個常見的詞彙而大致理解簡單的句子與問題。			8							
能聽懂簡單的問題。				6						
能聽懂簡短對話並掌握其中的關鍵訊息。							7			
能理解簡短對話的內容，並掌握對話的主旨。								8		
對於簡短、簡單的宣佈或說明，能聽懂其中的要點。									5	
能在聆聽日常生活中常見並簡短的宣布、說明、和語音媒體訊息時，掌握其主旨及重點資訊。										10

表 9 兒童華語文閱讀測驗雙向細目表

能力描述	題型	看圖辨義		圖文 連連看		閱讀材料			短文理解	
	測驗等級	萌芽	萌芽	成長	萌芽	成長	茁壯	成長	茁壯	
能理解常見、簡單的詞彙。		12	4							
能利用關鍵詞彙理解句子。				8						
能讀懂簡短材料中的短語，找到可預知的基本資訊。					4					
能讀懂簡短材料中的短語，如車票、便條、明信片上的文字及簡單的行走路線。						8				
能理解具日常功能性的簡單閱讀材料，並掌握其中的主旨與關鍵資訊。							12			
在視覺輔助下，能理解簡單簡短敘述中的要點。									9	
能理解簡單的故事與敘事短文，理解主要情節。										13

2. 建構效度

(1) 試題分析

兒童測驗組卷方式是依據 IRT 而來。IRT 的一項重要假設為：單向度假設。所謂單向度假設，指的是測驗中所有題目皆測量相同潛在特質，若忽略受測者回答試題是否仰賴單一特質這項資訊時，所獲得的試題參數及受測者能力估計值將會存在偏誤。

本節將採用 Winsteps 3.68.2 版的 Rasch 模式(如公式 1 所示)分析資料，以提供兒童測驗建構效度證據，結果如下所述。

由表 10 與表 11 可發現，萌芽級聽力理解測驗，試題的平均難度與考生能力估計值分別為-4.65 與-1.42，兩者差距 3.23 logits；閱讀測驗方面，試題的平均難度與受測者能力估計值則分別為-4.61 與-3.03，兩者差距 1.58 logits。顯示萌芽級測驗中，相較於聽力理解測驗，閱讀測驗之受測者平均能力與試題平均難度之間絕對值差異較小。

表 10 萌芽級試題難度估計分布

測驗內容	聽力	閱讀
平均值	-4.65	-4.61
標準差	0.71	0.52
最大值	-3.04	-3.55
最小值	-5.87	-5.70

表 11 萌芽級受測者能力估計分布

測驗內容	聽力	閱讀
平均值	-1.42	-3.03
標準差	1.53	1.45
最大值	-0.20	-0.29
最小值	-5.87	-8.92

成長級聽力與閱讀測驗結果(如表 12 與表 13)類似於萌芽級，試題的平均難度估計值與受測者能力估計值分別為聽力-3.19 與-0.28 以及閱讀-3.23 與-2.27，兩者差距各約為 2.91 logits 與 0.96 logit，閱讀測驗之受測者平均能力與試題平均難度之間絕對值差異較小。

表 12 成長級試題難度估計分布

測驗內容	聽力	閱讀
平均值	-3.19	-3.23
標準差	0.79	0.78
最大值	-1.61	-1.38
最小值	-4.54	-4.86

表 13 成長級受測者能力估計分布

測驗內容	聽力	閱讀
平均值	-0.28	-2.27
標準差	1.57	1.19
最大值	1.51	1.49
最小值	-6.65	-5.90

同樣的結果，在茁壯級(如表 14 及表 15)也可發現，聽力理解測驗中受測者平均能力與試題平均難度之絕對值差異比閱讀理解測驗大，兩者分別約為 2.18 logits 與 0.48 logit。

表 14 茁壯級試題難度估計分布

測驗內容	聽力	閱讀
平均值	-1.38	-2.07
標準差	0.67	0.64
最大值	-0.04	-0.21
最小值	-2.34	-2.87

表 15 茁壯級受測者能力估計分布

測驗內容	聽力	閱讀
平均值	0.80	-1.59
標準差	1.53	1.31
最大值	3.25	2.59
最小值	-3.53	-5.40

整體來說，兒童測驗本年度聽力與閱讀測驗各等級的考生平均能力估計值皆高於試題難度的平均值。造成此一現象可能的原因為，本測驗萌芽級、成長級、茁壯級的建議報考學時分別為 150 小時、300 小時與 450 小時；然而萌芽級、成長級和茁壯級有效填答樣本中，學習時數超過 600 小時的比例分別為 59.2%、

67.4%與 91.1%。由此可見，有相當高比例的考生學習時數遠超過建議報考時數，因此使得整體考生的能力值高於所報考等級的平均難度。

此外，各等級閱讀測驗受測者平均能力與試題平均難度之間絕對值差異比聽力測驗小，造成此現象的可能原因為參加本測驗的兒童多屬華裔，各等級有效填答樣本中，皆有半數左右或超過半數的考生表示經常或有時候與家人說中文，因此學習者在學習過程中可大量依靠聽和說的方式學習華語，此一現象反映於受測者聽力平均能力在三個等級皆有偏高的趨勢(見表 11、表 13 及表 15)；相較之下，在非華語地區，學習者大多數僅能透過學校課程及課外讀物學習中文字詞，能接觸到的視覺接收活動較少，因此，此一現象反映於受測者的閱讀平均能力與試題平均難度差異較小。同樣的現象可由圖 8 至圖 10 受測者與試題分布圖(Person - Item Map)觀察得知，將受測者能力估計值與試題難度估計值分別置於圖的左側和右側，由上而下將受測者能力由高至低排列、試題難度由難至易排列，可發現與聽力理解測驗相較，閱讀理解測驗試題的難度涵蓋範圍與受測者能力分布範圍間的差距較小，各等級聽力與閱讀理解測驗試題平均難度 95%信賴區間涵蓋的受測者能力分布分別為，萌芽級 17%與 37%、成長級 19%與 74%以及茁壯級 31%與 73%。

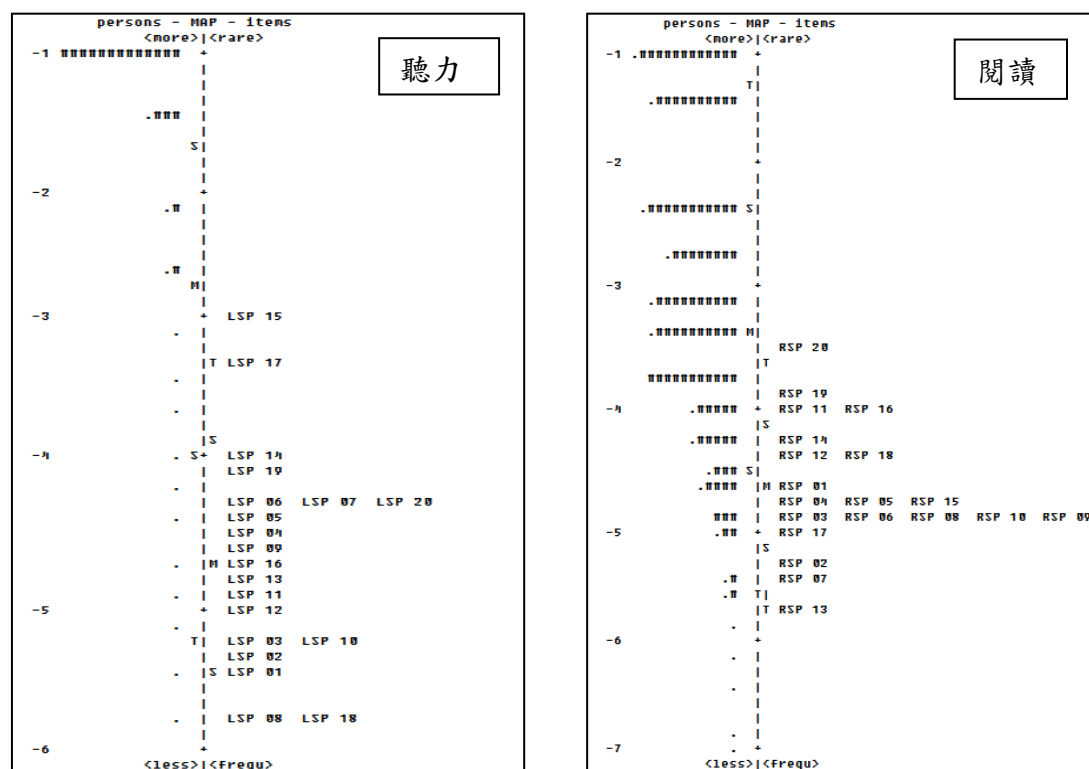


圖 8 萌芽級測驗受測者與題目分布圖

註：“#”在聽力與閱讀理解測驗分別表示 35 位及 8 位受測者；“.”表示一位受測者；LSP 01~LSP 20 和 RSP 01~RSP 20 表示試題編號。

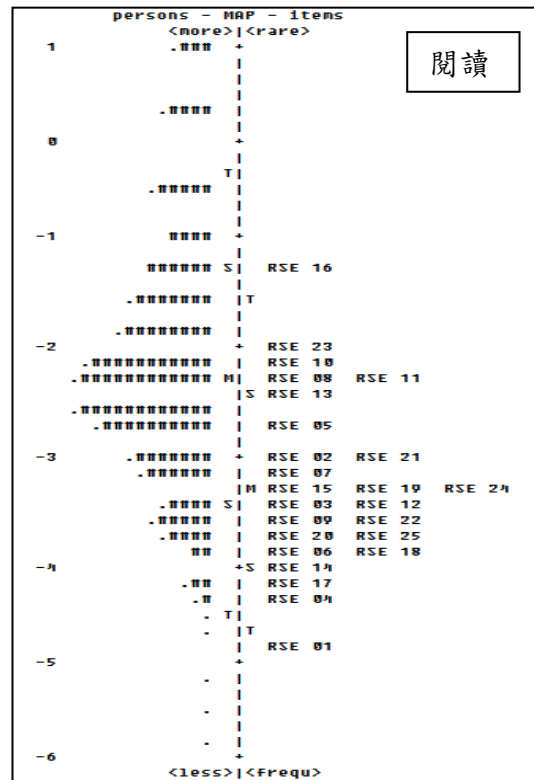
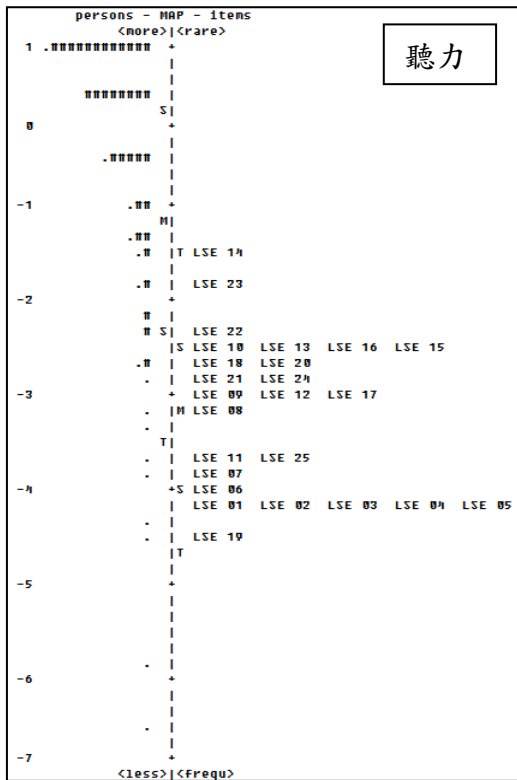


圖 9 成長級測驗受測者與題目分布圖

註：“#”在聽力與閱讀理解測驗分別表示 21 位及 7 位受測者；“.”表示一位受測者；LSE 01~LSE 25 和 RSE 01~RSE 25 表示試題編號。

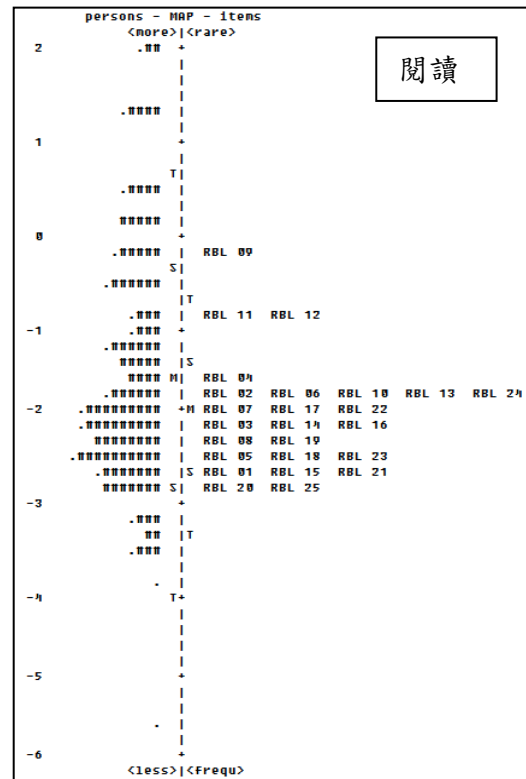
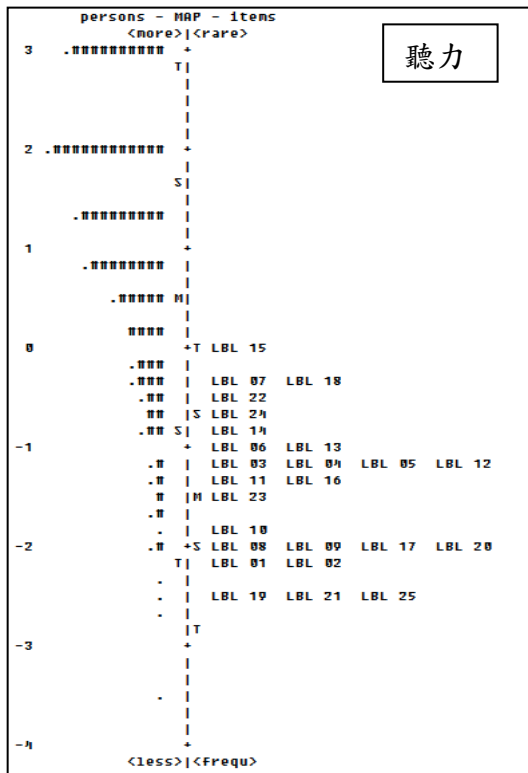


圖 10 茁壯級測驗受測者與題目分布圖

註：“#”在聽力與閱讀理解測驗分別表示 8 位及 5 位受測者；“.”表示一位受測者；LBL 01~LBL 25 和 RBL 01~RBL 25 表示試題編號。

由於預試時考生答題動機可能不若正式考試強烈，加上正式考試考生整體能力未必與試題難度相當，考生整體能力多半高於試題難度，適配度容易有偏低的現象，因此在正式考試採取較為寬鬆的標準，在試題品質分析方面，採用統計指標 Infit MNSQ 介於 0.6 到 1.4 以及 Infit ZSTD 介於-3 到 3 的標準評估試題與單向度 IRT 模式適配的程度。結果如表 16 顯示，萌芽級聽力與閱讀測驗 20 道試題中各有三道，成長級聽力與閱讀測驗 25 道試題中分別有七道、三道，以及茁壯級聽力測驗 25 道試題中有三道試題，茁壯級閱讀測驗 25 道試題中沒有試題不符合單向度 IRT 模式的假設。各等級在聽力與閱讀測驗適配率分別為，萌芽級 85% 與 85%、成長級 72% 與 88% 以及茁壯級 88% 與 100%。

一般來說，造成試題與模式不適配的原因有以下幾種，一為當受測者答題所需要的能力超過單一向度時，則可能造成試題不適配，因在 Rasch 模式的假設下，受測者答對試題機率僅受到單向度能力(此為兒童聽力或閱讀能力)所影響。二是當試題具有其他影響答題機率的特性，例如，試題的猜測力或鑑別度，而未能被 Rasch 模式納入解釋時，也會發生試題不適配情形。三是當受測者因為粗心或其他原因使得作答未符合其能力表現時，亦會造成試題估計不符合模式。例如，在模式假定受測者能力越高答對機率越高的條件下，高能力之受測者答錯較簡單題目時。

由於本測驗正式卷的試題皆經過預試，且獲得穩定難度參數估計值後所組成，故測驗觀察資料與測量模式不適配的原因應為第三種情形；在分析時，試題難度被設定為已知再進行模式適配度考驗，而受測者因粗心或其他原因使得作答未符合其能力表現，造成試題估計不符合模式。

表 16 試題適配分布

測驗等級	萌芽級		成長級		茁壯級	
	聽力	閱讀	聽力	閱讀	聽力	閱讀
正式考試						
總試題數	20	20	25	25	25	25
適配題數	17	17	18	22	22	25
不適配題數	3	3	7	3	3	0
試題適配率 (%)	85	85	72	88	88	100

以下進一步針對各等級測驗不適配試題進行選項分析，結果如表 17 至表 19 所示。表 17 至表 19 列出各等級測驗中不適配試題之難度參數估計值及估計誤

差、各選項選答人數、選答該選項的受測者平均能力估計值、平均能力估計誤差及點二系列相關係數。由表 17 可發現，萌芽級聽力測驗和閱讀測驗共六道不適配試題，選答正確選項人數都是最多的，正答選項平均能力估計值也是最高的，表示選答該選項的受測者平均能力高於選答錯誤選項者，如模式所預期，能力越高的受測者答對試題的機率越高。所有試題中正確選項的點二系列數值皆符合標準⁴，表示具有幅合效度(convergent validity)，即能力越高者答對試題的機率越高；錯誤選項值較低且幾乎為負值，表示具有區辨效度(discriminate validity)，可有效區分能力高與低的受測者(Linacre, 2010)。

表 17 萌芽級測驗不適配試題選項描述性摘要表

測驗類別	試題編號	難度	估計誤差	選項	計分	人數	平均能力	平均能力估計誤差	點二系列相關
聽力	LSP_02	-5.243	0.165	B	0	24	-3.81	0.24	-.26
				C	0	48	-3.59	0.17	-.31
				A	1	768	-1.19	0.05	.41
				M	0	4	-3.89	0.63	-.11
	LSP_03	-5.195	0.163	B	0	14	-4.55	0.19	-.29
				A	0	6	-4.43	0.24	-.18
				C	1	823	-1.34	0.05	.35
	LSP_09	-4.620	0.141	M	0	1	-2.75	--	.00
				A	0	12	-4.50	0.18	-.27
				C	0	19	-4.19	0.26	-.30
				B	1	812	-1.30	0.05	.40
	RSP_09	-4.893	0.096	M	0	1	-4.89	--	-.10
O				0	226	-3.84	0.06	-.28	
X				1	589	-2.61	0.06	.28	
M				0	29	-5.13	0.20	-.32	
閱讀	RSP_13	-5.697	0.117	B	0	7	-5.01	0.21	-.15
				C	0	9	-4.96	0.20	-.16
				A	0	10	-4.86	0.25	-.15
				D	0	8	-4.28	0.26	-.07
				E	1	802	-2.93	0.05	.27
				M	0	8	-5.53	0.67	-.17
RSP_17	-5.023	0.098	A	0	49	-4.04	0.15	-.16	
			B	0	210	-3.90	0.07	-.27	
			C	1	574	-2.59	0.06	.34	
			M	0	11	-4.99	0.48	-.16	

註：M 表示漏答；"--" 表人數為 1 人，無法計算平均能力估計標準差。

⁴ 點二系列相關係數標準為在 .30 以上，惟極端值難度試題，如低於 -4.0 或高於 -0.5，則標準放寬至 .15 以上。

表 18 為成長級聽力、閱讀測驗選項分析的結果，與萌芽級相同，十題不適配試題中選答正確選項人數都是最多的，該選項平均能力估計值也是最高的；至於點二系列相關數值，除 RSE_15 外，其餘九題正確選項的數值均符合標準，說明成長級中，在數據上呈現為不適配的這幾道試題仍具幅合效度與區辨效度，可有效區分能力高與低的受測者。

RSE_15 此次模式適配結果，Infit MNSQ 與 Infit ZSTD 分別為 1.54 與 9.90，檢視預試時的分析結果，兩項指標數值分別為 0.90 與 -1.61，符合模式適配標準。研發人員將持續觀察此題於正式考試後的分析結果，若未來於正式卷中仍持續發生不適配的情形，則考慮停止使用。

表 18 成長級測驗不適配試題選項描述性摘要表

測驗類別	試題編號	難度	估計誤差	選項	計分	人數	平均能力	平均能力估計誤差	點二系列相關
聽力	LSE_02	-4.169	0.166	B	0	6	-3.72	0.26	-.26
				A	0	21	-3.11	0.13	-.36
				C	1	819	-0.16	0.05	.44
				M	0	6	-3.86	0.83	-.23
	LSE_03	-4.165	0.166	C	0	32	-2.67	0.17	-.34
				A	0	38	-2.17	0.19	-.26
				B	1	770	-0.06	0.05	.43
				M	0	12	-2.03	0.69	-.13
	LSE_04	-4.134	0.164	C	0	37	-2.30	0.17	-.28
				B	0	52	-2.17	0.14	-.29
				A	1	757	-0.02	0.05	.41
				M	0	6	-4.03	0.74	-.25
	LSE_07	-3.809	0.149	B	0	50	-2.12	0.17	-.29
				C	0	60	-1.98	0.15	-.28
				A	1	737	0.01	0.05	.41
				M	0	5	-4.37	0.80	-.26
	LSE_20	-2.758	0.115	D	0	3	-3.31	0.49	-.15
				B	0	14	-2.69	0.22	-.22
				A	0	142	-1.57	0.11	-.26
				C	1	686	0.07	0.05	.34
	LSE_24	-2.928	0.119	M	0	7	-2.44	0.87	-.13
				B	0	127	-1.64	0.11	-.27
				A	0	94	-1.38	0.13	-.14
				C	1	629	0.17	0.06	.32
LSE_25	-3.664	0.142	M	0	2	-4.11	1.78	-.15	
			A	0	52	-2.17	0.15	-.29	
			C	0	52	-2.15	0.12	-.26	
			B	1	742	0.00	0.05	.40	
RSE_06	-3.888	0.089	M	0	6	-2.02	0.85	-.08	
			A	0	11	-4.00	0.14	-.20	
			B	0	48	-3.56	0.10	-.28	
			D	0	10	-3.45	0.20	-.11	
			C	0	19	-3.41	0.17	-.15	
			E	1	762	-2.11	0.04	.40	
RSE_14	-4.051	0.093	M	0	2	-3.80	0.29	-.07	
			C	0	168	-2.80	0.06	-.18	
			A	0	208	-2.62	0.05	-.09	
			B	1	455	-1.86	0.06	.23	
			M	0	21	-3.19	0.18	-.12	
			C	0	179	-2.72	0.05	-.17	
RSE_15	-3.195	0.079	A	0	320	-2.58	0.05	-.13	
			B	1	322	-1.61	0.08	.27	
			M	0	31	-3.21	0.15	-.16	
			C	0	31	-3.21	0.15	-.16	

註：M 表示漏答。

表 19 為茁壯級測驗的選項分析結果，所有不適配試題選答正確選項人數都是最多的，該選項平均能力估計值也是最高的，所有試題正確選項的點二系列數值亦符合標準，表示試題具有幅合效度；錯誤選項的點二系列值較低且幾乎為負值，表示試題具有區辨效度，可有效區分能力高與低的受測者。

表 19 茁壯級測驗不適配試題選項描述性摘要表

測驗類別	試題編號	難度	估計誤差	選項	計分	人數	平均能力	平均能力估計誤差	點二系列相關
聽力	LBL_09	-1.948	0.142	B	0	71	-0.63	0.15	-0.32
				A	0	59	-0.56	0.16	-0.27
				C	1	462	1.19	0.06	.45
				M	0	1	-1.29	--	-0.06
	LBL_17	-2.025	0.145	C	0	63	-0.84	0.15	-0.36
				A	0	31	0.16	0.24	-0.04
				B	1	498	1.04	0.06	.33
				M	0	1	0.14	--	.00
	LBL_24	-0.745	0.112	C	0	143	-0.17	0.09	-0.24
				B	0	70	-0.09	0.16	-0.16
				A	1	380	1.32	0.07	.32

註：M 表示漏答；“-”表人數為 1 人，無法計算平均能力估計標準差。

(2) 驗證性因素分析

除了透過試題分析來評估兒童華語文能力測驗是否具有建構效度之外，本測驗亦從結構方程模型(structural equation modeling)之驗證性因素分析來評估測驗的建構效度。由於兒童聽力和閱讀測驗題型之答題反應皆為非對即錯的二分名義變項，故使用可分析二元變項和類別變項的 Mplus 分析軟體(Muthén & Muthén, 2012)，來檢視萌芽級、成長級與茁壯級兒童聽力與閱讀測驗各題型所欲測得之能力，是否可經由受測者於試題的答題反應測知。每一等級之因素分析模型中，測量變數為每道試題，而潛在變數則為測驗題型，例如，本年度萌芽級兒童聽力測驗中，測量變數為 20 道試題，潛在變數為「聽句子選圖」、「看圖回答」與「聽句子連連看」。

此節使用前述之 Mplus 7.0 版進行類別變項之驗證性因素分析(CFA with categorical factor indicators)，估計方法採用「平均數與變異數修正後的加權最小平方值法」(weighted least squares means and variance adjusted, WLSMV)，各測

驗等級之驗證性因素分析結果則分別透過基本適配度及整體適配度指標進行模式評估。

依據Bagozzi和Yi(1988)以及Hu和Bentler(1998)，訂定出基本適配度的評估標準如下：(1)因素負荷量介於.50至.95之間；(2)相關係數不可大於1.0；(3)不能有過大的標準誤。至於整體適配指標，則採用卡方自由度比(χ^2/df)來評估整個模式與觀察資料的適配程度；以平方概似平方誤根係數(root mean square error of approximation；簡稱RMSEA)指標來評估整體模式的絕對適配度；以非規範適配指標(non-normed fit index；簡稱NNFI，亦稱為TLI)與比較適配指標(comparative-fit index；簡稱CFI)二項指標來評估整體模式增值適配度。判斷標準分別為： $\chi^2/df < 3$ 、RMSEA $< .08$ 、CFI和NNFI $> .90$ 。

基本適配指標部分，萌芽級、成長級與茁壯級兒童聽力與閱讀測驗正式考試驗證性因素分析結果如下圖11至圖13所示，由於篇幅有限，各等級各題因素負荷量等參數請見附件2至4。聽力測驗方面，萌芽級三因素模型驗證性因素分析結果顯示，除了LSP_17負荷量較高外(.99)，其餘各題因素負荷量介於.63至.90之間，平均值為.76；因素負荷量標準誤為0.03至0.06；因素負荷量統計考驗均達顯著水準($p < .05$)；潛在變項間的相關係數分別為.76、.76與1.01。成長級四因素模型驗證性因素分析結果顯示，各題因素負荷量介於.48至.91之間，平均值為.73；因素負荷量標準誤為0.02至0.10；因素負荷量統計考驗均達顯著水準($p < .05$)；各潛在變項間的相關係數介於.73至.98之間。至於茁壯級三因素模型驗證性因素分析結果顯示，各題因素負荷量介於.40至.82之間，平均值為.65；因素負荷量標準誤為0.04至0.06；因素負荷量統計考驗均達顯著水準($p < .05$)；各潛在變項間的相關係數分別為.98、.93、.99。

閱讀測驗方面，萌芽級三因素模型驗證性因素分析結果顯示，各題因素負荷量介於.47至.91之間，平均值為.69；因素負荷量標準誤為0.02至0.05；因素負荷量統計考驗均達顯著水準($p < .05$)；潛在變項間的相關係數分別為.57、.59與.61。成長級三因素模型驗證性因素分析結果顯示，各題因素負荷量介於.13至.86之間，平均值為.60；因素負荷量標準誤為0.03至0.06；因素負荷量統計考驗均達顯著水準($p < .05$)；潛在變項間的相關係數分別為.61、.48、.67。至於茁壯級二因素模型驗證性因素分析結果顯示，各題因素負荷量介於.32至.95之間，平均值為.57；因素負荷量標準誤為0.03至0.06；因素負荷量統計考驗均達顯著水準

($p < .05$)；潛在變項間的相關係數為.81。

根據以上指標分析，可知兒童聽力測驗與閱讀測驗各等級正式考試之驗證性因素模型基本適配度檢驗大致良好，試題因素負荷量介於.50至.95的比例在64%至96%之間，表示大多數試題均符合標準。若以一般測驗實務上可接受之因素負荷量.30為基準，兒童聽力測驗三個等級正式考試因素負荷量低於.30的試題比例均為0%，兒童閱讀測驗萌芽級與茁壯級低於.30的試題比例亦為0%，成長級有一題低於.30。顯示僅極少數的試題因素負荷量較低，而因素負荷量較低之試題，於未來組正式卷時將特別留意。

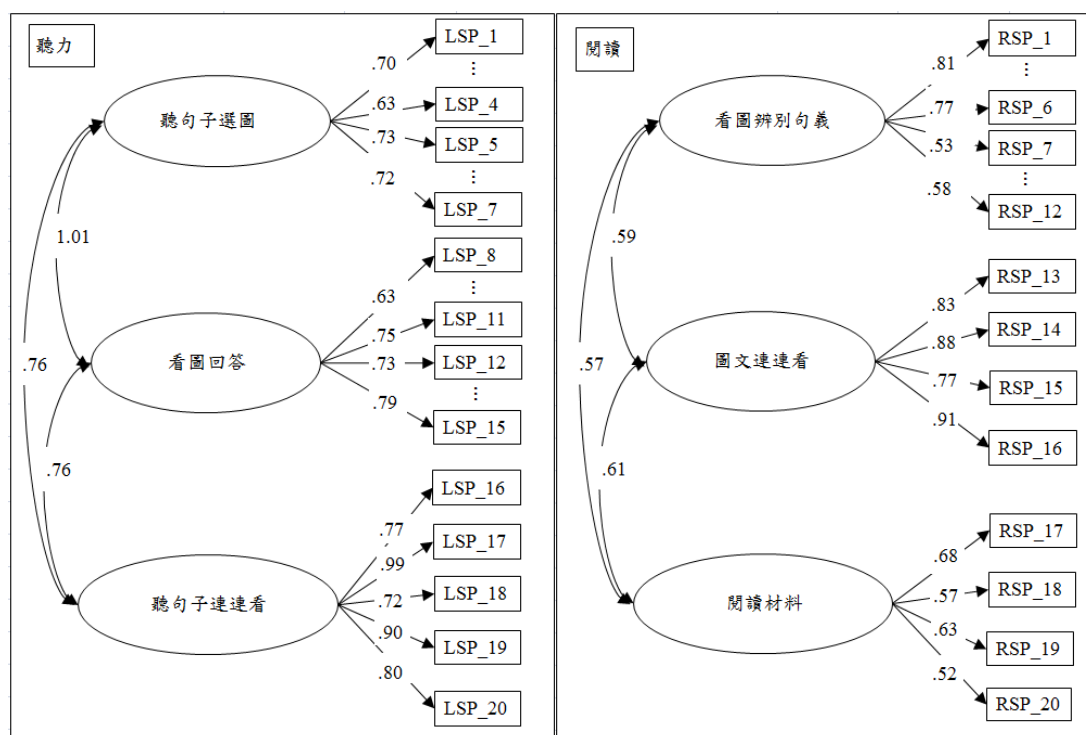


圖 11 萌芽級兒童測驗驗證性因素分析

註：LSP_1-LSP_7所屬題型為「聽句子選圖」；LSP_8-LSP_15所屬題型為「看圖回答」；LSP_16-LSP_20所屬題型為「聽句子連連看」。RSP_1-RSP_12所屬題型為「看圖辨別句義」；RSP_13-RSP_16所屬題型為「圖文連連看」；RSP_17-RSP_20所屬題型為「閱讀材料」。

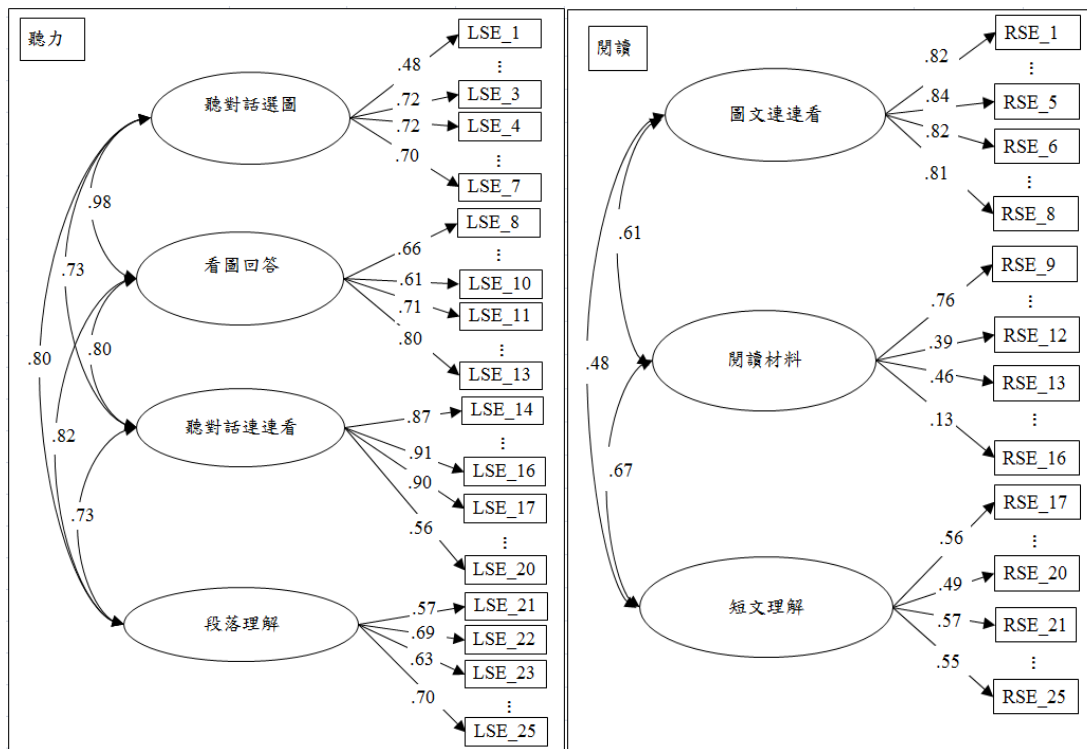


圖 12 成長級兒童測驗驗證性因素分析

註：LSE_1-LSE_7所屬題型為「聽對話選圖」；LSE_8-LSE_13所屬題型為「看圖回答」；LSE_14-LSE_20所屬題型為「聽對話連連看」LSE_21-LSE_25所屬題型為「段落理解」。RSE_1-RSE_8所屬題型為「圖文連連看」；RSE_9-RSE_16所屬題型為「閱讀材料」；RSE_17-RSE_25所屬題型為「短文理解」。

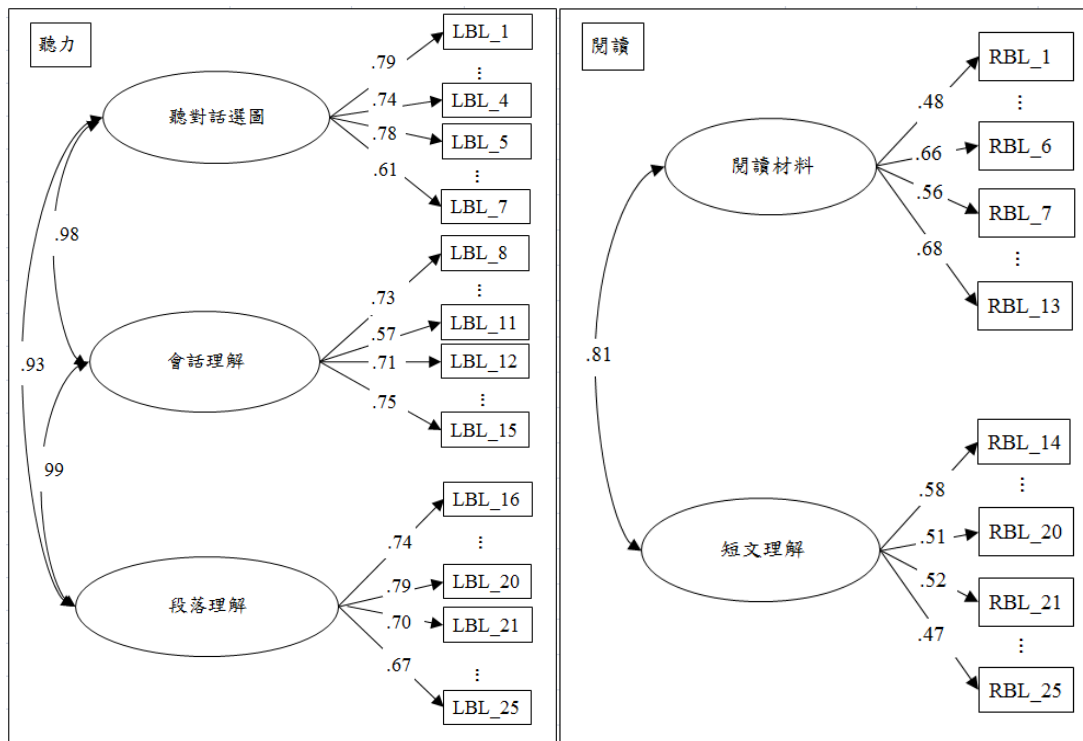


圖 13 茁壯級兒童測驗驗證性因素分析

註：LBL_1-LBL_7所屬題型為「聽對話選圖」；LBL_8-LBL_15所屬題型為「會話理解」；LBL_16-LBL_25所屬題型為「段落理解」。RBL_1-RBL_13所屬題型為「閱讀材料」；RBL_14-RBL_25所屬題型為「短文理解」。

整體模式適配度主要在評鑑整個模式與觀察資料的適合程度，相當於模式的外在品質。首先，經由卡方自由度比(χ^2/df)來評估模式適配度，其原因為卡方分配易受到自由度與估計參數數量的影響，當自由度越大或所需估計的參數個數越多時，卡方值越大，造成假設模式適配度不佳的可能性越大，此時卡方值越不能反映理論模式與觀察資料的適配程度。由表 20 可發現，聽力測驗方面，兒童測驗各等級正式考試的卡方自由度比均小於 3，表示模式適配度良好(Wheaton, Muthen, Alwin & Summers, 1977)。而在絕對適配度評估上，各等級分析結果，RMSEA 數值均小於.08，表示三種模式皆符合絕對適配度指標。增值適配度評估部分，各等級分析結果，CFI 和 NNFI 數值皆大於.90，顯示三模式均符合增值適配度指標。閱讀測驗方面，萌芽級與成長級測驗正式考試的卡方自由度比略大於 3，茁壯級測驗小於 3，萌芽級與成長級模式適配度較不理想。在絕對適配度評估上，各等級分析結果，RMSEA 數值均小於或等於.08，表示三種模式皆符合絕對適配度指標。增值適配度評估部分，各等級分析結果，CFI 和 NNFI 數值皆大於或接近.90，顯示三模式大致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，兒童聽力和閱讀測驗各等級聽力測驗大體上具有建構效度，各測驗試題分別可測得不同的聽力或閱讀理解能力，而各題型間的相關係數達到中度至高度相關，反映測得一致之聽力或閱讀理解能力。

表 20 整體模式適配度指標摘要表

測驗類別	檢驗模式	χ^2	χ^2/df	RMSEA	CFI	NNFI
聽力測驗	萌芽級三因素模式	225.82	1.35	.02	.99	.99
	成長級四因素模式	465.04	1.73	.03	.97	.97
	茁壯級三因素模式	343.92	1.26	.02	.99	.98
閱讀測驗	萌芽級三因素模式	1012.37	6.06	.08	.87	.85
	成長級三因素模式	1158.22	4.26	.06	.85	.83
	茁壯級三因素模式	595.21	2.17	.04	.93	.93

五、 結論

本文為 2014 年兒童華語文能力測驗技術報告，闡述內容共包含兩個部分，第一部分分別針對兒童華語文能力測驗之能力描述、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。第二部分則為 2014 年度之兒童測驗整體性測驗信度與效度評估，目的在檢視其是否能夠發揮測驗效用並確切地測量受測者的目標潛在能力。

在測驗信度分析方面，本文採內部一致性指標觀察測驗試題間之相關性，藉以確認整份測驗中的試題是否皆測量到相同潛在特質及其程度為何。此外，也透過測驗訊息量，檢視測驗是否對通過門檻提供較高的訊息量，有較佳的測量精準度。在測驗效度分析部分，本測驗在組卷完成後，即通過專家審查、評估試題於不同等級中的適切性以確保試題品質，針對試題內容的審核步驟完整且明確，可賴以確保測驗之內容效度。在此應補充說明的是，由於本測驗試題內容的適切性主要仰賴專家審查，唯有通過審查並完成預試的試題方能使用於正式考試中施測，故專家審查程序是重要且不可忽略的內容效度證據。為了提供更嚴謹、全面及量化的專家效度證據，本會未來在邀請外審專家審查時，除目前所提供的各等級能力描述與各大題所測能力等資訊外，也考慮提供更具針對性的試題檢核項目資料，如設計審題問卷，請專家針對每一道試題進行詳細評分或判斷，通過此一辦法獲得專家對於試題評估的量化結果。

除了具備測驗之內容效度證據之外，在施測完成後，研發人員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，與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。總體而言，本文之信度、效度分析結果顯示，本會於 2014 年度所舉辦之全球性兒童測驗正式考試，所獲得之受測者成績及結果已相當可靠，並可確實測量到測驗研發初期所訂定之目標能力。

六、文獻

- 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。
- 國家華語測驗推動工作委員會(2015a)。華語文能力測驗技術報告 2013(1)：聽力測驗信效度(ISBN：978-986-92167-2-2)。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(2015b)。華語文能力測驗技術報告 2013(1)：閱讀測驗信效度(ISBN：978-986-92167-3-9)。新北市：國家華語測驗推動工作委員會。
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16(1), 74-94.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Dorans, N. J. (2000). Scaling and scales. In Wainer, H (Eds.), *Computerized Adaptive Testing: A Primer* (pp. 135-158). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hu, L.T., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424 – 453.
- Linacre, J.M. (2009). Winsteps (Version 3.68.2) [Computer Software]. Beaverton, Oregon: Winsteps.com.
- Linacre, J. M. (2010). Winsteps Rasch measurement computer program User's Guide. Beaverton, Oregon: Winsteps.com
- McKay, P. (2006). *Assessing Yong Language Learners*. Cambridge: Cambridge University Press.
- Muthén, L.K. and Muthén, B.O. (2012). Mplus (Version 7.0) [Computer Software]. Los Angeles, CA: Muthén & Muthén.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* Copenhagen: Institute of Educational Research. (Expanded edition, 1980. Chicago: The University of Chicago Press.)
- Wheaton, B., Muthen, B., Alwin, D., & Summers, G. (1977). Assessing the reliability and stability in panel models. In D.R. Heise (ed.), *Sociological Methodology*. San Francisco: Jossey-Bass.

附件 1 兒童測驗測驗各等級正式考試成績單樣張

附件 1-1 萌芽級測驗正式考試成績單樣張



兒童華語文能力測驗成績單

Children's Chinese Competency Certification Test Score Report

姓名 Name : 王大衛 David Wang
 報考等級 Test Level : 萌芽級 (Sprouting)
 考生編號 Test No. : 1405226160401001
 考試日期 Test Date : 2014 年 5 月 12 日 (May 12, 2014)
 考試地區 Test Area : 韓國 (Korea)

聽力測驗成績 The Score of Listening Comprehension	閱讀測驗成績 The Score of Reading Comprehension	總成績 Total Score
100 / 100	77 / 100	177 / 200


測驗結果 Test Result	通過 (Pass)
-------------------------	------------------

總成績達 117 分者，即通過萌芽級測驗。
 Test takers require an overall score of at least 117 to pass the Sprouting Test.


每項描述下最高可達四顆星(★)
 Test takers can receive four stars (★) at most on each description.

聽力能力描述 Listening Proficiency Description		
一	能聽懂簡單常見的詞彙。 Can understand simple and familiar spoken words and phrases.	★★★★
二	能聽懂簡單的問題。 Can understand simple spoken questions.	★★★★
三	能聽懂非常緩慢、發音清楚、簡短、簡單的句子。 Can understand short, simple sentences when spoken slowly and clearly.	★★★★

閱讀能力描述 Reading Proficiency Description		
一	能理解基本的詞語及簡單的句子。 Can understand basic words and simple sentences.	★★★★
二	在有視覺輔助的前提下，藉由掌握基本詞語能大致理解簡單的書寫材料。 Can mostly understand simple written materials through mastery of basic words, with the help of visual aids.	★★



國家華語測驗推動工作委員會
Steering Committee for the Test Of Proficiency-Huayu



附件 1-2 成長級測驗正式考試成績單樣張



兒童華語文能力測驗成績單

Children's Chinese Competency Certification Test Score Report

姓名 Name : 王大衛 David Wang
 報考等級 Test Level : 成長級 (Seedling)
 考生編號 Test No. : 1405226170401001
 考試日期 Test Date : 2014 年 4 月 2 日 (April 2, 2014)
 考試地區 Test Area : 美國 (U.S.A.)

聽力測驗成績 The Score of Listening Comprehension	閱讀測驗成績 The Score of Reading Comprehension	總成績 Total Score
100 / 100	52 / 100	152 / 200

測驗結果 Test Result	通過 (Pass)
-------------------------	------------------

總成績達 116 分者，即通過成長級測驗。
 Test takers require an overall score of at least 116 to pass the Seedling Test.

每項描述下最高可達四顆星(★)
 Test takers can receive four stars (★) at most on each description.

聽力能力描述 Listening Proficiency Description		
一	能聽懂有關日常生活的問題。 Can understand questions related to daily life.	★★★★
二	能聽懂非常緩慢、發音清楚、簡短、簡單的對話。 Can understand very slow, clear, and simple short dialogues.	★★★★
三	對於簡短、簡單的宣佈或說明，能聽懂其中的要點，例如：自我介紹、電話留言、口頭通知、學校或課堂上的宣佈、天氣預報、注意事項或指令等。 Can understand the main points when listening to short and simple announcements and instructions, such as introductions, phone messages, notifications, school and classroom announcements, and weather reports.	★★★★

閱讀能力描述 Reading Proficiency Description		
一	能掌握句子中的關鍵詞彙，並理解句義。 Can fully grasp the key words in a sentence and understand its meaning.	★★★★
二	能理解簡短材料中的基本資訊或說明。 Can read and understand basic information or instructions from short written materials.	★★
三	在視覺輔助下，能理解簡單、簡短的敘事短文中的要點。 Can understand the main points of simple and short narrative passages, with the help of visual aids.	★



國家華語測驗推動工作委員會
 Steering Committee for the Test Of Proficiency-Huayu



附件 1-3 茁壯級測驗正式考試成績單樣張



兒童華語文能力測驗成績單

Children's Chinese Competency Certification Test Score Report

姓名 Name：王~~大~~衛 **David Wang**
 報考等級 Test Level：茁壯級 (Blossoming)
 考生編號 Test No.：1405226180401001
 考試日期 Test Date：2014 年 3 月 22 日 (March 22, 2014)
 考試地區 Test Area：德國 (Germany)

聽力測驗成績 The Score of Listening Comprehension	閱讀測驗成績 The Score of Reading Comprehension	總成績 Total Score
77 / 100	66 / 100	143 / 200

測驗結果 Test Result	通過 (Pass)
-------------------------	------------------

總成績達 118 分者，即通過茁壯級測驗。
 Test takers require an overall score of at least 118 to pass the Blossoming Test.

每項描述下最高可達四顆星(★)
 Test takers can receive four stars (★) at most on each description.

聽力能力描述 Listening Proficiency Description		
一	能聽懂簡短對話並掌握其中關鍵訊息。 Can understand simple dialogues and fully grasp essential points.	★★★
二	能聽懂簡短對話的內容，並掌握對話的主旨。 Can understand the content of short dialogues and fully grasp their gist.	★★★★
三	聆聽以可預知的日常生活為主題，簡單、簡短且常見的宣布、說明和語音媒體錄音時，能掌握其主旨及重要資訊。 Can understand the gist and main points when listening to short, simple announcements, instructions and recordings related to everyday life.	★★★★

閱讀能力描述 Reading Proficiency Description		
一	能理解具日常功能性的簡單閱讀材料，並掌握其中的主旨、要點與可預知的特定資訊。 Can understand basic reading materials about everyday activities, and can fully grasp the gist, main points and specific predictable information.	★★★
二	能理解與自身經驗或日常生活相關的簡單故事與敘事短文中的主要情節。 Can understand the plots of simple stories and narrative passages related to personal experience or everyday life.	★★★★



國家華語測驗推動工作委員會
 Steering Committee for the Test Of Proficiency-Huayu



附件 2 萌芽級兒童測驗各題因素負荷量及殘差變異量摘要表

測驗類別	題型	試題編號	因素負荷量	標準誤	P 值
聽力	聽句子選圖	LSP_1	.70	0.04	.00
		LSP_2	.69	0.05	.00
		LSP_3	.75	0.05	.00
		LSP_4	.63	0.05	.00
		LSP_5	.73	0.04	.00
		LSP_6	.81	0.04	.00
		LSP_7	.72	0.04	.00
	看圖回答	LSP_8	.63	0.06	.00
		LSP_9	.76	0.05	.00
		LSP_10	.82	0.03	.00
		LSP_11	.75	0.04	.00
		LSP_12	.73	0.04	.00
		LSP_13	.76	0.04	.00
		LSP_14	.83	0.03	.00
	聽句子連連看	LSP_15	.79	0.04	.00
		LSP_16	.77	0.04	.00
		LSP_17	.99	0.03	.00
		LSP_18	.72	0.05	.00
		LSP_19	.90	0.03	.00
		LSP_20	.80	0.03	.00
閱讀	看圖辨別句義	RSP_1	.81	0.02	.00
		RSP_2	.87	0.02	.00
		RSP_3	.77	0.03	.00
		RSP_4	.84	0.02	.00
		RSP_5	.79	0.03	.00
		RSP_6	.77	0.03	.00
		RSP_7	.53	0.04	.00
		RSP_8	.49	0.04	.00
		RSP_9	.47	0.04	.00
		RSP_10	.49	0.04	.00
		RSP_11	.56	0.04	.00
		RSP_12	.58	0.04	.00
	圖文連連看	RSP_13	.83	0.04	.00
		RSP_14	.88	0.03	.00
		RSP_15	.77	0.04	.00
		RSP_16	.91	0.03	.00
	閱讀材料	RSP_17	.68	0.05	.00
		RSP_18	.57	0.05	.00
		RSP_19	.63	0.05	.00
		RSP_20	.52	0.05	.00

附件 3 成長級兒童測驗各題因素負荷量及殘差變異量摘要表

測驗類別	題型	試題編號	因素負荷量	標準誤	P 值
聽力	聽對話選圖	LSE_1	.48	0.10	.00
		LSE_2	.91	0.04	.00
		LSE_3	.72	0.05	.00
		LSE_4	.72	0.05	.00
		LSE_5	.67	0.06	.00
		LSE_6	.82	0.04	.00
		LSE_7	.70	0.05	.00
	看圖回答	LSE_8	.66	0.06	.00
		LSE_9	.70	0.05	.00
		LSE_10	.61	0.05	.00
		LSE_11	.71	0.05	.00
		LSE_12	.84	0.03	.00
	聽對話連連看	LSE_13	.80	0.03	.00
		LSE_14	.87	0.03	.00
		LSE_15	.85	0.03	.00
		LSE_16	.91	0.02	.00
		LSE_17	.90	0.03	.00
		LSE_18	.91	0.03	.00
	段落理解	LSE_19	.80	0.04	.00
		LSE_20	.56	0.05	.00
		LSE_21	.57	0.06	.00
		LSE_22	.69	0.05	.00
		LSE_23	.63	0.05	.00
		LSE_24	.55	0.05	.00
		LSE_25	.70	0.05	.00
閱讀	圖文連連看	RSE_1	.82	0.03	.00
		RSE_2	.75	0.03	.00
		RSE_3	.86	0.03	.00
		RSE_4	.79	0.03	.00
		RSE_5	.84	0.03	.00
		RSE_6	.82	0.03	.00
		RSE_7	.78	0.03	.00
		RSE_8	.81	0.03	.00
		RSE_9	.76	0.04	.00
	閱讀材料	RSE_10	.60	0.05	.00
		RSE_11	.39	0.05	.00
		RSE_12	.39	0.05	.00
		RSE_13	.46	0.05	.00
		RSE_14	.36	0.05	.00
		RSE_15	.47	0.05	.00
		RSE_16	.13	0.06	.04
	短文理解	RSE_17	.56	0.05	.00
		RSE_18	.61	0.04	.00
		RSE_19	.44	0.05	.00
		RSE_20	.49	0.05	.00
		RSE_21	.57	0.04	.00
		RSE_22	.71	0.04	.00
		RSE_23	.45	0.05	.00
		RSE_24	.50	0.04	.00
		RSE_25	.55	0.04	.00

附件 4 茁壯級兒童測驗各題因素負荷量及殘差變異量摘要表

測驗類別	題型	試題編號	因素負荷量	標準誤	P 值
聽力	聽對話選圖	LBL_1	.79	0.05	.00
		LBL_2	.72	0.06	.00
		LBL_3	.68	0.06	.00
		LBL_4	.74	0.04	.00
		LBL_5	.78	0.04	.00
		LBL_6	.48	0.06	.00
		LBL_7	.61	0.05	.00
	會話理解	LBL_8	.73	0.04	.00
		LBL_9	.64	0.05	.00
		LBL_10	.82	0.04	.00
		LBL_11	.57	0.05	.00
		LBL_12	.71	0.05	.00
		LBL_13	.55	0.05	.00
		LBL_14	.45	0.06	.00
		LBL_15	.75	0.04	.00
	段落理解	LBL_16	.74	0.05	.00
		LBL_17	.52	0.06	.00
		LBL_18	.76	0.04	.00
		LBL_19	.68	0.05	.00
		LBL_20	.79	0.04	.00
		LBL_21	.70	0.05	.00
		LBL_22	.49	0.05	.00
		LBL_23	.40	0.06	.00
		LBL_24	.47	0.05	.00
		LBL_25	.67	0.05	.00
閱讀	閱讀材料	RBL_1	.48	0.05	.00
		RBL_2	.63	0.04	.00
		RBL_3	.59	0.04	.00
		RBL_4	.59	0.04	.00
		RBL_5	.55	0.04	.00
		RBL_6	.66	0.04	.00
		RBL_7	.56	0.05	.00
		RBL_8	.78	0.03	.00
		RBL_9	.58	0.05	.00
		RBL_10	.68	0.04	.00
		RBL_11	.95	0.03	.00
		RBL_12	.34	0.05	.00
		RBL_13	.68	0.04	.00
	短文理解	RBL_14	.58	0.04	.00
		RBL_15	.59	0.05	.00
		RBL_16	.64	0.04	.00
		RBL_17	.45	0.05	.00
		RBL_18	.63	0.04	.00
		RBL_19	.44	0.05	.00
		RBL_20	.51	0.05	.00
		RBL_21	.52	0.05	.00
		RBL_22	.43	0.05	.00
		RBL_23	.48	0.05	.00
		RBL_24	.32	0.06	.00
		RBL_25	.47	0.05	.00

書 名：兒童華語文能力測驗技術報告—2014(5)

聽力、閱讀測驗信效度

出 版 者：國家華語測驗推動工作委員會

24449 新北市林口區仁愛路一段 2 號

886-2-7734-5638

印 刷 者：上校文化印刷有限公司

80744 高雄市三民區通化街 88 巷 26 號

886-7-311-6011

出版日期：2016 年 11 月

定 價：新台幣 100 元

版權所有

翻印必究

