

新版「華語文能力測驗」效標關聯效度研究

The Research Of Criterion-related Validity Of Updated Test Of Proficiency-Huayu

張可家

國家華語測驗推動工作委員會

kejia@sc-top.org.tw

摘要

本研究旨在探討新版「華語文能力測驗」的效標關聯效度。華測會於 2009 年 7 月與國立臺灣師範大學國語中心合作，在學期末舉行密集班及中高級班「華語文能力測驗」施測，透過問卷調查方式請華語教師評估華語學習者聽力與閱讀兩部份的語言能力，再分別和華語學習者的聽力和閱讀兩部份測驗成績加以比對，探討華語學習者的語言能力表現與華語文能力測驗成績之間的關聯性。針對此研究目的，本研究欲探討以下問題：華語教師評估語言能力表現較佳的學習者，是否在測驗成績的表現上也較佳？比對學習者能力等級與測驗平均數後，發現華語教師評估語言能力表現較好的學習者，測驗成績也較好。為進一步檢驗新版「華語文能力測驗」的效度，於是又將教師評估學生能力等級結果與測驗成績進行斯皮爾曼等級相關分析(Spearman rank correlation)，結果亦顯示兩者間存在顯著水準。研究結果顯示，新版「華語文能力測驗」能有效區辨華語學習者聽力和閱讀兩部份語言能力高低程度的差異。

關鍵字：新版「華語文能力測驗」、效標關聯效度、斯皮爾曼等級相關

一、 研究背景

近年，國際間著名的語言能力考試紛紛對應「歐洲語言學習、教學、評量共同參考架構」(The Common European Framework of Reference for Languages: Learning, Teaching, Assessment, 簡稱 CEF 或 CEFR)，例如：多益測驗(TOEIC)、托福測驗(TOEFL)、劍橋國際英語認證(Cambridge Main Suite)、劍橋職場英語檢測(BULATS)、德語檢定考試(TestDaF)、日本語能力考試、法語檢定考試(DELF)、全民英檢(GEPT)等等，都依據 CEFR 能力指標，將考試等級重新規劃。

台灣教育部於 2005 年公佈的「推動英語能力檢定測驗處理原則」公文中指出各機關院校在推動英語能力檢定測驗時，除其他法令另有規定外，應參考 CEF 架構，依測驗對象、目的及需求選擇適宜之測驗工具，並避免採用單一或少數測驗工具。除了英語能力的檢定外，教育部亦積極在華語文能力方面推動 CEF 架構，例如：於 2007 年邀請法國漢語總督學白樂桑造訪台灣進行學術交流。

「國家華語測驗推動工作委員會」(以下簡稱華測會)亦深切理解推動語言能力指標之重要性，為與國際語言能力標準接軌，成為提供華語學習者衡量其語言能力的國際評量工具。華測會於 2008 年起進行華語文聽讀能力測驗的改版計畫，參考 CEF 語言能力架構各等級能力描述，調整「華語文能力測驗」(Test Of Proficiency-Huayu)初等、中等及高等測驗題型及架構。基礎級測驗由於設計之初即參考 CEF 的 A2 等級進行試題研發，故未做調整，新版「華語文能力測驗」三個等級為「進階級」、「高階級」與「流利級」測驗，分別對應 CEF 語言能力架構的 B1、B2 以及 C1 等級。

以往關於「華語文能力測驗」的效度相關研究，主要以內容效度來評估測驗之效度，分析每份試卷內容題數、比重分配以及主題分佈是否恰當，然而，對於一個具有公信力和影響力的測驗，必須不斷地在效度研究上持續進行驗證，成為提供考生評量語言能力的國際性測驗評量工具。因此，華測會於 2009 年 7 月與國立臺灣師範大學國語中心合作，利用學期末舉行密集班及中高級班華語文能力測驗的機會，並透過華語教師評估華語學習者語言能力，整理評估結果與華語學習者測驗成績之間的關聯性，以作為新版「華語文能力測驗」的效度依據。

二、 研究目的

效度是測驗中最重要的特徵之一。效度指測驗分數的正確性，也就是說測驗能夠測量到它所想要測量特質的程度（郭生玉，2000），更具體地說，就是測驗能夠達到某種目的的程度（Mehrens & Lehmanmm, 1978）。在應用上，因為測驗目的不同，所使用的效度估計方法也不同，美國心理學會將效度分為三種類型：內容效度（content validity）、效標關聯效度（criterion-related validity）和構念效度（construct validity）。本研究為效標關聯效度研究，效標關聯效度是以經驗性的方法，研究測驗分數與一些外在效標間的關係（Thorndike & Hagen, 1977）。

在進行此類研究前，一般的作法是找出一個具公信力、能夠互相參照的測驗，而且這兩個測驗之間應具有相近的結構，評量的是相似的能力。目前在華語測驗領域中，還無法找到等級規劃同樣對應 CEF 語言能力架構的測驗，做為此次的外在效標。有學者（Bachman & Palmer, 1989；Oscarson, 1989）提出語言能力評量表是一種評估語言能力有效、可靠的方式。全民英檢（語言測驗訓練中心，2005）和中國漢語水平考試（王小玲，2006）也曾使用過語言能力評量表來評估語言學習者語言能力。由此可見，語言能力評量表的確是一種能夠有效評估語言學習者語言能力的工具。

本研究目的在檢驗新版「華語文能力測驗」的效標關聯效度，透過華語教師評估華語學習者聽力與閱讀兩部份的語言能力，再和華語學習者測驗成績加以比對，探討華語學習者的語言能力表現與華語文能力測驗成績之間的關聯性。針對上述研究目的，本研究欲探討以下問題：

華語教師評估語言能力表現較佳的學習者，是否在測驗成績的表現上也較佳？

第三節針對 1. 研究對象；2. 研究工具；3. 研究步驟；4. 資料分析，分別說明研究方法。

三、 研究方法與分析

I. 研究對象

研究對象為 2009 年 7 月國立台灣師範大學國語教學中心密集班、中高級班學生，總計 61 個班級，476 名學生，共 45 位華語教師參加此次研究。研究者根據學生學習情況建議報考等級，但學生仍可自行選擇報考的測驗等級。

II. 研究工具

1. 新版「華語文能力測驗」

新版「華語文能力測驗」為一種外語/第二語言能力測驗，測驗對象為母語非華語之人士，新版「華語文能力測驗」參照 CEF 架構各等級能力描述，對「華語文能力測驗」初等、中等及高等測驗題型及架構做調整。基礎級測驗由於設計之初即參考 CEF 架構 A2 等級進行試題研發，故未做調整。調整後新版「華語文能力測驗」分別為「進階級」、「高階級」與「流利級」測驗，分別對應 CEF 架構的 B1、B2 以及 C1 等級。

此次研究所使用測驗等級分別為基礎級、進階級、高階級與流利級預試試卷，各等測驗均分為聽力和閱讀二個部份。基礎級測驗共有 80 題，聽力與閱讀測驗各有 40 題；進階、高階和流利級測驗分別有 100 題，聽力和閱讀測驗各占 50 題。計分方式皆為一題一分，答錯不倒扣。

2. 語言能力評量表

發給華語教師填寫的語言能力評量表內容參考「歐洲語言評量手冊」(European Language Portfolio, ELP)，此評量手冊乃根據 CEF 架構所設計，依據不同語言能力等級，分別列出各等級語言學習者能做到何種表現的語言能力描述，提供語言學習者和語言教師隨時檢驗、掌握語言能力學習情況。

研究者編擬語言能力評量表之初，先從「歐洲語言評量手冊」語言能力描述中，挑選出符合華語文特性的能力描述及新版「華語文能力測驗」的測驗題型可測得的語言能力，以供此次研究使用。另外，考量若問卷內容過多，將會降低填寫者填寫意願，於是配合華語學習者報考的測驗等級，將評量表內容分為 A2、B1 與 B2 三種版本，均以中文呈現，每一種版本皆包含三個等級的語言能力描述，並區分成聽力及閱讀二大部分，例如：A2 版本評量表，包含 A1、A2、B1 三個等級的語言能力描述，其餘皆依此類推。各等

級評量表之題數分佈，見表 1。

華語教師填寫評量表時，需逐一評估班上學生的能力，根據學生聽力和閱讀兩部份的語言能力表現進行圈選評量表，若學生能達到評量表中所列出的語言能力描述，則圈選「Y」；若目前無法完成，則圈選「N」。

表 1 A2、B1 及 B2 版本語言評量表題數分佈

版本	能力	語言能力等級						合計
		A1	A2	B1	B2	C1	C2	
A2	聽力	3	8	6			34	
	閱讀	4	8	5				
B1	聽力		5	9	3		35	
	閱讀		6	9	3			
B2	聽力			7	6	4	36	
	閱讀			8	6	5		

III. 研究步驟

華測會於 2009 年 7 月與國立臺灣師範大學國語中心合作，在學期末舉行密集班及中高級班華語文能力測驗，鼓勵學生參與，但不具強迫性，由研究者提供學生建議報考等級，但學生仍可自行選擇測驗等級報名參加考試。學生報名後，研究者根據學生報考之測驗等級，於測驗實施後，以班級為單位發放語言能力評量表問卷請任課教師填寫。報考基礎級測驗考生發給 A2 版本評量表；報考進階級測驗考生發給 B1 版本評量表；報考高階與流利級測驗發給 B2 版本評量表。評量表填寫時間為二週，二週後回收評量表進行分析。

IV. 資料分析

評量表回收後，首先分別統計聽力及閱讀部分的能力描述圈選數量，進行語言能力等級的評估。若教師勾選學生語言能力的題數達到某一等級勾選數量的通過門檻¹，則視為具備該等級能力。原則上，學生必須先通過低一級的標準，才能獲得更高的等級。

以 B1 版本評量表為例，任課華語教師填寫的評量表內容包含 A2、B1、B2 等級的語言能力描述。若在聽力部份的 A2 等級勾選題數為 3 題(含)以上，則視為具有 A2 等級，接著再進一步檢視較高等級的勾選情形，若在 B1 等級勾選題數為 7 題，而 B2 等級勾選

¹ 通過標準為 75%的勾選率，但若無法整除為 0.75，則微幅降低通過門檻。

題數為 1 題，則確定為 B1 等級。

但若 A2 等級勾選 3 題，B1 等級勾選 5 題，而 B2 等級勾選題數為 3 題，因為已達到 A2 與 B2 等級通過門檻，雖未通過 B1 等級門檻，此一情形仍歸類為 B1 等級。若未通過 A2 等級門檻，但通過 B1 以及 B2 等級的門檻，亦歸類為 B1 等級。

假使學生未通過每一版本評量表的最低等級，因無法確切得知能力等級的下限，則以「-」表示，如 A2 版本即為「A1-」表示、B1 版本為「A2-」表示。各版本評量表之題數及通過門檻、語言能力等級評估情形，如下表 2 至表 5 所示。

表 2 A2 版本評量表題數及通過門檻

	聽力能力			閱讀能力		
能力等級	A1	A2	B1	A1	A2	B1
評量表題數	3	8	6	4	8	5
通過門檻	2	6	4	3	6	3

表 3 B1 版本評量表題數及通過門檻

	聽力能力			閱讀能力		
能力等級	A2	B1	B2	A2	B1	B2
評量表題數	5	9	3	6	9	3
通過門檻	3	6	2	4	6	2

表 4 B2 版本評量表題數及通過門檻

	聽力能力			閱讀能力		
能力等級	B1	B2	C1	B1	B2	C1
評量表題數	7	6	4	8	6	5
通過門檻	5	4	3	6	4	3

表 5 語言能力等級評估情形(以 B1 評量表為例)

A2 等級	B1 等級	B2 等級	能力等級 分類結果
通過	通過	通過	B2
通過	通過	未通過	B1
通過	未通過	通過	B1
未通過	通過	通過	B1
通過	未通過	未通過	A2
未通過	未通過	通過	A2-
未通過	通過	未通過	A2-
未通過	未通過	未通過	A2-

標記教師評估的學生聽力及閱讀語言能力等級後，接著再和學生華語文能力測驗成績進行比對。第四節將針對 1. 問卷回收情形；2. 能力等級與測驗表現：測驗平均數比較；3. 能力等級與測驗表現：斯皮爾曼等級相關分析；4. 結論與建議，分別探討考生聽力、閱讀語言能力表現與測驗成績之間的相關性。

四、研究結果與討論

I. 問卷回收情形

此次研究，共發出 496 份問卷，回收 407 份，回收率為 82.1%。A2 版本評量表共回收 34 份，B1 版本評量表共回收 98 份，B2 版本評量表共回收 275 份。評量表回收後檢查填寫情形，若漏答超過百分之十以上則視為無效樣本，聽力和閱讀部份分開計算。去除無效樣本後，A2 版本聽力評量表共 31 份、閱讀共 31 份，B1 版本聽力評量表共 90 份、閱讀共 93 份，B2 版本聽力評量表共 248 份、聽力共 247 份，各版本有效樣本數分佈情形，見表 6。

表 6 評量表回收情形

評量表版本	回收	聽力有效樣本數	閱讀有效樣本數
A2 版	34	31	31
B1 版	98	90	93
B2 版	275	248	247

II. 能力等級與測驗表現：測驗平均數比較

本小節將比對教師評估學生語言能力等級結果和測驗成績平均數表現，探討學生聽力、閱讀語言能力的表現。以下依照學生參加測驗等級分別說明，共四個測驗等級：1. 參加基礎級測驗學生表現；2. 參加進階級測驗學生表現；3. 參加高階級測驗學生表現；4. 參加流利級測驗學生表現。

參加基礎級測驗學生表現：

聽力部份，學生能力等級與聽力測驗平均成績比對後發現，能力等級為 A2 的學生，聽力測驗成績平均數為 34 分，被評為 B1 等級的學生，聽力測驗成績平均數為 35.48 分。教師評估學生聽力能力等級與測驗平均成績之詳細數據，見表 7、圖 1。由此可見，學生聽力能力等級越高，測驗成績的平均得分也越高。

閱讀部份，學生能力等級與閱讀測驗平均成績比對後可以發現，教師評估能力等級為 A2 的學生，閱讀測驗成績平均數為 33.17 分，被評為 B1 等級的學生，閱讀測驗成績平均數為 34.75 分，教師評估學生閱讀能力等級和測驗平均成績之詳細數據，見表 8 與圖 2。由此可見，學生閱讀能力等級越高，學生的測驗平均得分也越高，惟成績增加的幅度不若聽力表現明顯。

表 7 聽力等級與基礎級測驗表現

	等級	人數	平均數	標準差
教師評估	A1	1	26.00	—
	A2	4	34.00	1.826
	B1	21	35.48	2.272
	總計	26	34.88	2.847

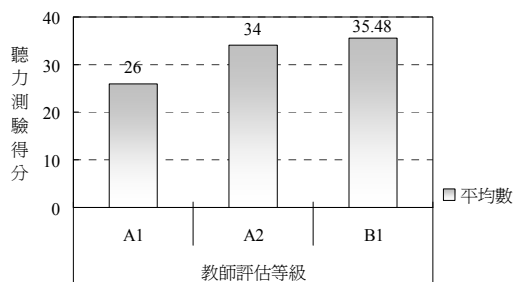


圖 1 聽力等級之基礎級聽力測驗長條圖

表 8 閱讀等級與基礎級測驗表現

	等級	人數	平均數	標準差
教師評估	A1	2	32.00	1.414
	A2	12	33.17	4.970
	B1	12	34.75	3.019
	總計	26	33.81	3.980

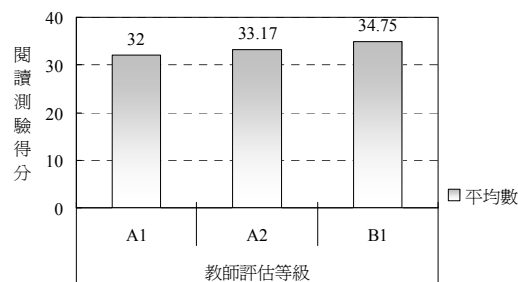


圖 2 閱讀等級之基礎級聽力測驗長條圖

參加進階級測驗學生表現：

聽力部份，教師評估學生能力等級與聽力測驗平均成績比對後可以發現，能力等級為 A2 等級的學生，聽力測驗平均數為 36.4 分，被評為 B1 等級的學生，測驗平均數為 42.21 分，被評為 B2 等級的學生，測驗平均數為 42.58 分，呈現出測驗平均數隨著能力等級提高而隨之增加的趨勢，詳細相關數據，見表 9、圖 3。由此可見，教師評估學生聽力能力等級越高，學生測驗成績的平均得分也隨著提高。

閱讀部份，教師評估學生能力等級與閱讀測驗成績平均比對後可以發現，能力等級為 A2 的學生，測驗平均成績為 42.66 分，被評為 B1 等級的學生，測驗平均數為 43.1，從教師評估閱讀能力等級與測驗表現可以發現，學生的閱讀測驗平均數從 A1 到 B2 等級呈現出逐漸增加的趨勢，詳細數據，見表 10、圖 4。

由此可見，教師評估學生閱讀能力等級越高，學生測驗成績的平均得分也隨著提高。

表 9 聽力等級與進階級測驗表現

	等級	人數	平均數	標準差
教師 評估	A2-	2	28.00	5.657
	A2	15	36.40	4.388
	B1	14	42.21	3.926
	B2	24	42.58	7.541
	總計	55	40.27	6.827

表 10 閱讀等級與進階級測驗表現

	等級	人數	平均數	標準差
教師 評估	A2-	2	29.00	5.657
	A2	35	42.66	4.291
	B1	10	43.10	4.677
	B2	9	45.22	3.073
	總計	56	42.66	4.981

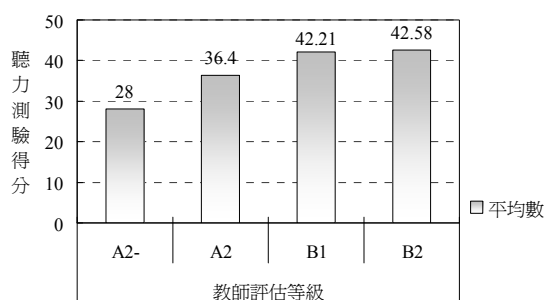


圖 3 聽力等級之進階級聽力測驗長條圖

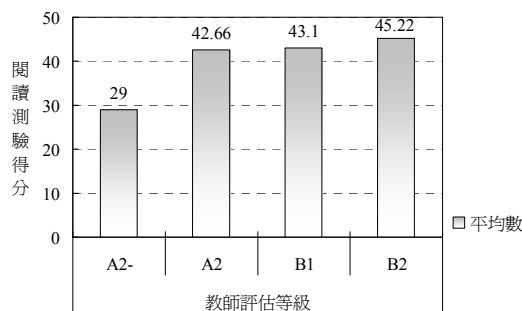


圖 4 閱讀等級之進階級聽力測驗長條圖

參加高階級測驗學生表現：

聽力部份，教師評估學生能力等級與聽力測驗平均成績比對後可以發現，能力等級為 B1 等級的學生，聽力測驗平均數為 33.46 分，被評為 B2 等級的學生，測驗平均數為 36.14 分，被評為 C1 等級的學生，測驗平均數為 43.08 分，詳細數據，見表 11、圖 5。

從教師評估聽力能力等級與測驗平均數可以觀察到，教師評估聽力能力等級越高的學生，其聽力測驗平均得分也隨著越高。

閱讀部份，教師評估學生能力等級與閱讀測驗平均成績比對後可以發現，教師評估學生能力等級為 B1 的學生，測驗平均成績為 26.41 分，被評為 B2 等級的學生，測驗平均數為 28.56 分，被評為 C1 等級的學生，測驗平均數為 32.33 分，詳細數據，見表 12、圖 6。由此可見，教師評估閱讀能力等級越高的學生，測驗成績平均數也越高。

表 11 聽力等級與高階級測驗表現

	等級	人數	平均數	標準差
教師評估	B1-	29	31.41	8.051
	B1	28	33.46	7.255
	B2	58	36.14	6.752
	C1	12	43.08	4.337
	總計	127	35.13	7.631

表 12 閱讀等級與高階級測驗表現

	等級	人數	平均數	標準差
教師評估	B1-	67	24.96	8.140
	B1	22	26.41	6.681
	B2	25	28.56	7.943
	C1	12	32.33	8.359
	總計	126	26.63	8.132

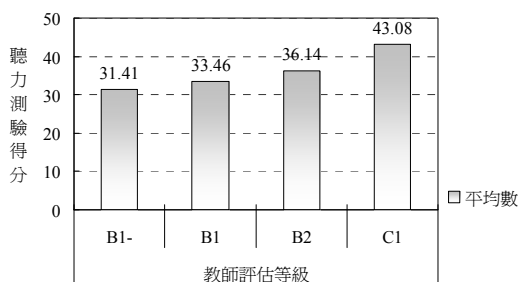


圖 5 聽力等級之高階級聽力測驗長條

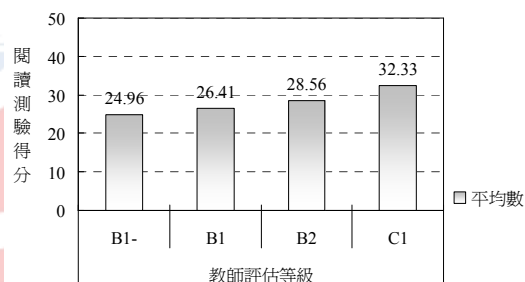


圖 6 閱讀等級之高階級聽力測驗長條圖

參加流利級測驗學生表現：

聽力部份，詳細數據，見表 13、圖 7，見教師評估學生能力等級與聽力測驗平均成績比對後可以發現，在教師評估學生能力等級為 B2 學生和被評為 C1 等級的學生測驗平均數雖然沒有呈現微幅增加的趨勢，但不難發現，被評為 B2 等級學生人數僅 1 人，因為樣本數過少，不具代表性，希望未來能提高樣本數，再提供數據以供參考。整體來看，從學生能力等級被評為 B1-到 C1 等級的學生，測驗成績平均數仍呈現增加的趨勢。

閱讀部份，教師評估學生能力等級與閱讀測驗平均成績比對後可以發現，教師評估學生能力等級為 B1 的學生，測驗平均成績為 17.50 分，被評為 B2 等級的學生，測驗平均數為 27.33 分，被評為 C1 等級的學生，測驗平均數為 33.5 分，詳細數據，見表 14、圖 8。由此可見，教師評估閱讀能力等級越高的學生，測驗成績平均數也隨之提高。

表 13 聽力等級與流利級測驗表現

	等級	人數	平均數	標準差
教師評估	B1-	8	29.75	7.106
	B1	5	30.20	12.950
	B2	1	36.00	—
	C1	6	35.50	9.854
	總計	20	31.90	9.335

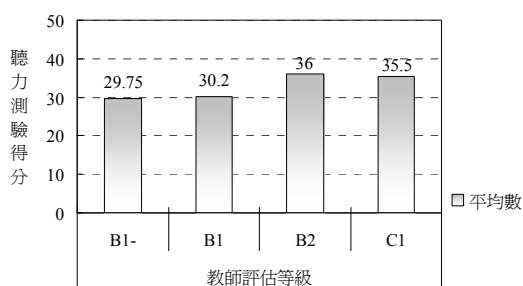


圖 7 聽力等級之流利級聽力測驗長條圖

表 14 閱讀等級與流利級測驗表現

	等級	人數	平均數	標準差
教師評估	B1-	8	17.50	5.318
	B1	2	17.50	0.707
	B2	3	27.33	11.015
	C1	8	33.50	9.411
	總計	21	25.00	10.531

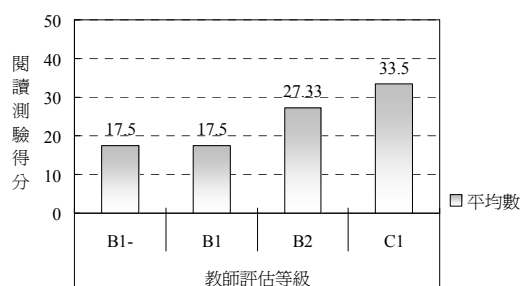


圖 8 閱讀等級之流利級聽力測驗長條圖

綜合上述結果發現，教師評估學生能力等級越高，學生測驗成績平均數也隨之提高，兩者之間的關係呈現出相當一致的情形，比對能力等級評估結果與測驗成績之間的結果後發現，學生測驗成績平均數越高的學生，教師評估學生能力等級也越高，由此可知，新版「華語文能力測驗」能有效的區分出學生語言能力的高低。

III. 能力等級與測驗表現：斯皮爾曼等級相關分析

為進一步檢驗新版「華語文能力測驗」是否能有效區辨考生能力高低程度的差異，也就是測驗的效度，於是，研究者進一步將教師評估學生能力等級結果，與測驗得分進行斯皮爾曼等級相關分析(Spearman rank correlation)，以求得新版「華語文能力測驗」的效標關聯效度，檢驗學生能力等級和測驗成績之間是否具有正相關？是否教師評估學生語言能力等級較高的學生，在測驗上的表現也較佳？

但由於參加基礎級和流利級測驗的學生人數較少，未達 30 人，若進行相關分析結果亦不具備代表性，因此，本研究僅針對樣本數較多的進階級和高階級測驗進行斯皮爾曼等級相關分析。分析資料時，分別將聽力和閱讀的能力等級結果與聽力成績和閱讀成績求等級相關。

參加進階級測驗的學生，在聽力測驗部份，相關分析結果發現，教師評估學生聽力

發表於 2009 年全美中文教師學會年會(ACTFL)，11 月 20-22 日，美國聖地牙哥

能力等級與聽力測驗成績的相關值為 0.553($p < 0.01$)，達到顯著水準；顯示教師評估聽力等級較高的學生，在聽力理解測驗的表現也較佳，有中等程度的正相關。在閱讀理解部份，教師評估學生閱讀能力等級與閱讀測驗成績的相關數值為 0.296($p < 0.05$)，同樣達到顯著水準，因此，由斯皮爾曼等級相關分析結果發現，教師評估學生能力等級，不論在聽力或閱讀能力上，與測驗成績均達到正相關。

參加高階級測驗的學生，在聽力測驗部份，相關分析結果發現，教師評估聽力能力等級與聽力理解測驗分數的相關值為 0.389($p < 0.01$)，達到顯著水準；顯示教師評估聽力等級較高的學生，在聽力測驗的表現也較佳。在閱讀理解部份，測驗成績與教師評估等級的相關值為 0.262($p < 0.01$)，亦達到顯著水準；顯示教師評估學生能力等級，不論是聽力或閱讀能力上，與測驗成績均達到正相關。

從斯皮爾曼等級相關分析結果發現，不論參加進階級測驗的學生或參加高階級的學生，教師評估學生聽力、閱讀能力等級與聽力、閱讀測驗成績之間均存在正相關。因此，由等級相關結果得知，新版「華語文能力測驗」能有效區分出學生能力等級的差異程度。

IV. 結論與建議

本研究透過教師評估學生語言能力等級與學生測驗成績相互比對後發現，不論在聽力或閱讀能力，教師評估學生語言能力等級越高，相對的學生的測驗成績平均數也隨著提高。再者，為了進一步檢驗新版「華語文能力測驗」是否能有效區辨考生能力高低程度的差異，於是又將教師評估學生能力等級結果與學生測驗成績進行斯皮爾曼等級相關分析(Spearman rank correlation)，研究結果顯示，參加進階級和高階級測驗的學生，不論在聽力或閱讀測驗部份，教師評估學生語言能力等級和測驗成績都達到顯著水準。

由於本次研究僅針對教師評估學生語言能力等級和測驗成績進行比對，建議未來研究能增加學生自評語言能力等級部份，將學生自評語言能力等級、教師評估語言能力等級和學生測驗成績三者交互進行比對，提供更多實證數據加以驗證。另外，本次研究也受限於參加基礎級和流利級測驗學習者樣本數較少，因此無法針對這兩個等級進一步提供斯皮爾曼等級相關數據，希望未來研究能針對這兩個測驗等級蒐集更多樣本進一步加以探討。本研究僅為新版「華語文能力測驗」效標關聯效度研究之開端，後續仍有待更深入持續的研究進行驗證，以期新版「華語文能力測驗」能更具公信力、更臻盡善盡美，並且提供華語學習者、華語教師以及教學機構評量華語文能力的有效工具。

參考文獻

1. 中文部分

王小玲 (2006)。HSK 初中等效度研究報告。《語言教學與研究》，第 6 期。

郭生玉 (2000)。《心理與教育測驗》。台北市：精華。

語言訓練測驗中心 (2005)。共同英語能力指標與全民英檢級數參照第一階段研究。取自

<http://www.lttc.ntu.edu.tw/academics/geptreport.htm>。

2. 英文部分

Bachman & Palmer (1989) The construct Validity of Self-rating of Communicative Language Ability, *Language Testing*, V6 N1:14-29.

Mehens, W. A. , & Lehmann, I. J. (1978) *Measurement and Education in Education and Psychology*. N. Y. : Holt, Rinehart and Winston.

Oscarson, M. (1989). Self-assessment of Language Proficiency : Rationale and Applications , *Language Testing*, V6 N1:1-13.

Thorndike, R. L., & Hagen, E. P. (1977) *Measurement and Education in Psychology and Education*.(4th ed.) N. Y.: John Wiley & Sons.