

華語學習者詞彙廣度與深度知識測試之 設計初探*

雷達莉 蕭惠貞
國立臺灣師範大學華語文教學系暨研究所

摘要

現今詞彙知識研究領域中，針對英語學習者詞彙知識之研究明顯居多，且多半採取由專家設計的標準化詞彙知識測試工具進行研究（Qian 2000, 2002；白麗梅與陳艷玲 2005；李曉 2007；王申 2010）。反觀華語教學方面，由於缺乏標準化的測試工具，華語詞彙知識研究的發展仍落後於英語相關研究。有鑑於此，本文以英語詞彙知識測試工具為參考基準，試圖設計出華語詞彙廣度及深度知識的測試工具，以填補華語詞彙知識研究的空缺。經過施測，發現本測試工具含以下優點：（一）具高信度（廣度、深度的折半信度皆高達 0.99）；（二）具一定程度的效度（因素分析從兩種測試工具各提取兩個被測量的隱性變項，而兩個變項在廣度、深度的總方差解釋分別為 72.39% 與 63.09%）；（三）能準確地估計學習者詞彙量（95% 信賴水準）；（四）能區分四個不同程度（初級、初中級、中級及中高級）學習者的詞彙知識。此外，兩種工具的整合設計使得測試結果具有描述廣度與深度關聯性之功能。

關鍵詞：詞彙知識 詞彙廣度 詞彙深度 詞彙測試設計 華語教學

* 本研究曾於 2017 年「第十四屆國際漢語教學學術研討會」上（The 14th ICCLP Conference）宣讀部分成果，並進行討論。感謝與會者對本研究的寶貴建議。另外，承蒙臺灣科技部計畫經費的挹注（MOST 105-2410-H-003-090），使本實驗得以順利完成。特別感謝研究助理梁安琪協助文章校訂。更感謝本期刊三位匿名審查委員所提出的修改寶貴建議與評論，使本文得以減少疏漏，更臻完善。文中有任何其他殘存問題，責在作者。

1. 前言

「四種傳統基本語言能力，包括閱讀、寫作、口說與聽力，皆奠基於學習者的詞彙知識¹」（Barrow et al. 1999: 223-224）。李曉（2007）探究英語詞彙知識與語言綜合能力之關係，發現詞彙廣度和深度知識均能有效預測語言綜合能力。此外，許多探討英語詞彙知識與閱讀理解關係的研究亦發現詞彙廣度、深度知識與閱讀理解存在著高度²正相關（Qian 2000, 2002；白麗梅與陳艷玲 2005；王申 2010）。值得注意的是，上述英語詞彙知識的相關研究皆採用標準化的詞彙知識測試予以施測，如廣度測試工具“Vocabulary Level Test”（以下簡稱 VLT）（Nation 1990），以及深度測試工具“Word Associates Test”（以下簡稱 WAT）（Read 2000）。由此可見，標準化測試工具之設計是詞彙知識研究發展極為重要的一環。

較之英語詞彙知識研究，針對華語詞彙知識的研究不但比較少見，更缺乏標準測試工具，以致各研究因受到不同測試工具的影響而取得不一致的結果。錢旭菁（2002）及辜玉旻等人（2012）曾設計華語詞彙知識測試工具，錢（2002）以英語 VLT 為基礎設計華語詞彙廣度知識測試工具，然本文認為該測試工具仍無法廣泛應用於其它詞彙廣度知識的相關研究。其原因如下：

（一）選定目標詞之前，未調查所使用的詞頻參考是否符合學習者詞彙習得的情況³；（二）部分測試內容採取翻譯形式⁴，較難針對不同母語背景的受測者施測，測試時不但要了解各受測者的母語背景外，更可能造成不一致的結果；（三）該工具未考察測試信度與效度，且受測者人數僅 18 位，不足以進一步測試信、效度；（四）設計範圍僅含約 3,000 個常用詞彙，故較無法準確測量高級學習者的詞彙量⁵。另一方面，辜等人（2012）的測試工具以漢

¹ 原文為“The four traditional basic language skills of reading, writing, speaking and listening are based upon the student’s vocabulary knowledge.” (Barrow et al. 1999: 223-224)。

² 於預測研究中，理想的相關係數須達 0.60 以上（王文科與王智弘 2014）。因此，本文定義 0.60 以上為中度相關性，0.70 以上為高度相關性。

³ 詳見 2.1 節「詞彙頻率」所述。

⁴ 測試要求受測者把目標詞翻譯成自己的母語。

⁵ 以漢語水平考試（Hanyu Shuiping Kaoshi，簡稱 HSK）為標準，五級學習者的詞彙量為 2,500 個詞彙，最高級（六級）學習者則能達 5,000 個詞彙。以華語文能力測驗（Test of Chinese as Foreign Language，簡稱 TOCFL）為標準，高階級（四級）學習者詞彙量已能達 5,000 個，精通級（六級）甚至能達到 8,000 個詞彙。

語幼童為設計對象，因此不適合用於成人學習者的研究。

綜上所述，詞彙知識是語言綜合能力的要件，然目前詞彙知識與語言能力相關之研究多以英語學習者為對象，此與英語詞彙知識測試工具發展較為健全有關。然目前華語學界尚未有這樣的標準化系統，有鑒於此，本研究主要目標乃試圖設計一套適合華語詞彙知識的標準測試工具，希冀能完善華語詞彙測試工具，並促進華語詞彙知識與語言能力關係之相關研究。

本文共分為五個段落。在前言之後，第二節為文獻探討，而第三節則是介紹研究材料與方法。第四節則進行研究結果討論與分析、修正建議、本研究限制與未來研究建議。最後，第五節總結本文之主要成果。

2. 文獻探討

以下我們將分別介紹詞彙頻率的重要性與詞彙知識的定義和內容。此外，我們也探究英文詞彙知識測驗較為常使用的測試工具之特色。

2.1 詞彙頻率

在語言學習過程中，某些詞語較常出現且較早學會，有些則相反。Milton (2009) 認為詞彙頻率的重要性即在於（一）預測哪些詞彙於語言學習過程中較為常見或常被聽到；（二）預測上述所提及的詞彙的出現機率。因此，詞頻可大致預測學習者於不同階段的詞彙學習進程。了解詞彙頻率與詞彙習得之關係後，才能設計出理想的詞彙知識測試工具。事實上，許多常用的英語詞彙知識測試工具即是以詞彙頻率為重要的設計基礎之一，此可證明詞頻的重要性。

然而，若以母語者語料庫的詞彙頻率作為設計學習者測試工具的基礎，須謹慎考察該語料庫的詞頻能否反映學習者詞彙習得的真實情況。換言之，必須確認語料庫的高頻詞彙為學習者較早學會者，反之亦然。據此，Meara (1992) 提出了語言學習者典型詞頻概貌 (Word Frequency Profile) 的概念 (如圖 1)。據圖 1 所示，1000 詞頻帶等級⁶越高，學習者學會的詞彙比率也越高，即學習者應該較易習得語料庫中屬於高頻率的詞彙。

資料來源：<http://www.sc-top.org.tw/chinese/LS/test5.php>。

⁶ 圖 1 中的 1000 詞頻帶等級包括第 1 到第 5 個等級。第 1 個 1000 詞頻帶是指排序 1~1000 的詞頻，第 2 個 1000 詞頻帶是指排序 1001~2000 的詞頻，如此類推。

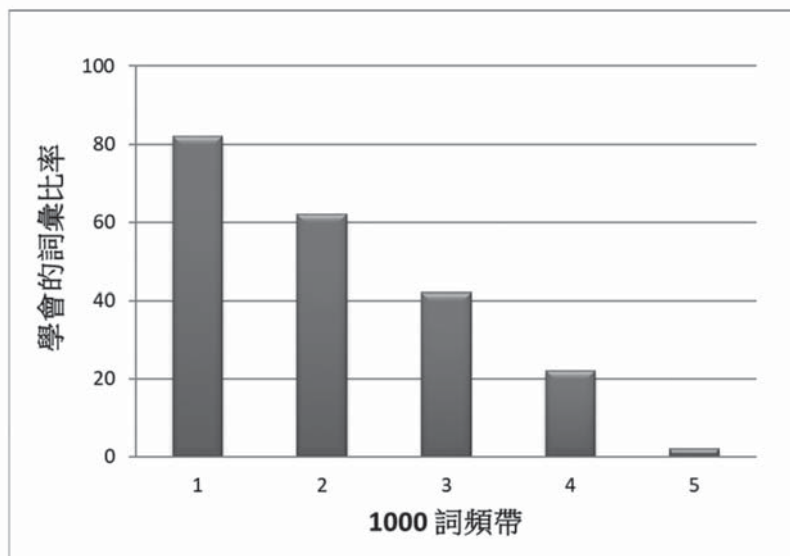


圖 1：語言學習者的典型詞頻概貌 (Meara 1992: 6)

2.2 詞彙知識

2.2.1 詞彙知識定義與組成

近年來，許多學者對於詞彙知識或「了解一個詞」提出若干定義。「了解一個詞」是認識該詞的口說、書寫形式及意義 (Nation 2001)，或認識並能正確使用該詞 (Pignot-Shahov 2012)。然而，Laufer 與 Goldstein (2004) 指出詞彙知識並非僅理解詞彙形式與意義之關聯，故簡略的定義實際上無法完整描述其複雜概念。另，Richards (1976) 提出「了解一個詞」包括了解詞語於口語或書面語出現的機率、在不同語境中的使用限制、句法行為、派生詞、詞彙網絡中的聯想詞、語義價值及多義詞等七個面向。以 Richards (1976) 的架構為基礎，Nation (1990) 進一步描述詞彙知識的組成，包括詞形（口說形式及書寫形式）、使用（語法規範及搭配方式）、功能（詞頻率及恰當性）、詞義（概念義及聯想詞）。然而，Meara (1996) 認為過於詳細的組成使測試工具需要大量的測試項目，因此將導致對於詞彙知識研究的施測困難，據此，其主張儘管詞彙概念具高度複雜性，詞彙知識仍應以少數且容易被測量的組成來描述。

Meara (1996) 所提出的詞彙知識架構僅包含詞彙量及詞彙組織兩個組成。詞彙量是指詞彙知識的基礎組成，在學習者詞彙量較少的情況下，它可

能是決定詞彙能力的唯一因素。然而，當學習者詞彙量達到某臨界點後，學會多少詞彙量的影響力將逐漸減弱。另一方面，詞彙組織基本上與 Richards (1976)「詞彙網絡中的聯想詞」或 Nation (1990)「聯想詞」具有相同概念，指學習者對於詞與詞之間關係的理解。隨著詞彙量對詞彙影響力的減弱，詞彙組織之影響力相對地增強。由此可見，Meara (1996) 詞彙知識架構的優點在於詞彙量及詞彙組織兩個組成，是反映整體詞彙網絡的屬性，並非單一詞彙的屬性。

2.2.2 詞彙廣度與深度知識

Qian (2000: 283) 認為詞彙廣度知識是指「學習者的詞彙量或學習者了解基本義項的詞彙數量⁷」，而詞彙深度知識是指「學習者對某個目標詞彙所具有的各方面知識之程度，亦即學習者對該詞彙的理解度⁸」。此外，Milton (2009: 13) 對於兩種詞彙知識的定義如下：「詞彙廣度知識是指學習者所知詞彙數量，詞彙深度知識則是指學習者對這些詞的了解⁹」。根據上述定義與 Meara (1996) 的詞彙知識概念，本文所界定的詞彙廣度知識對應 Meara (1996) 的詞彙量組成，詞彙深度知識則指詞彙組織組成。詞彙組織除包含聯想詞外，更涉及 Nation (1990) 的「搭配方式」，因目標詞與搭配詞亦屬於一種詞與詞的關係，即當學習者看到目標詞彙時也會聯想到其搭配詞。

2.3 本研究中詞彙知識測試工具設計的參考來源

2.3.1 英語詞彙廣度知識測試工具：Vocabulary Level Test (VLT)

詞彙水平測試 (Vocabulary Level Test, VLT) 為最常用的英語詞彙廣度知識測試工具。按照其所採用的詞頻統計，VLT 測試分成五部分，各代表不同詞頻帶詞彙之測試¹⁰，其中 2000 詞頻帶與 3000 詞頻帶均含高頻詞彙，學

⁷ 原文為“...vocabulary size, or the number of words for which a learner has at least some minimum knowledge of meaning.” (Qian 2000: 283)。

⁸ 原文為“a learner's level of knowledge of various aspects of a given word, or how well the learner knows this word.” (Qian 2000: 283)。

⁹ 原文為“Breadth of knowledge refers to the number of words a learner knows and depth of knowledge refers to what the learner knows about these words.” (Milton 2009: 13)。

¹⁰ 詞頻帶包括 2000 詞頻帶、3000 詞頻帶、5000 詞頻帶、學術詞彙表 (University Word List) 及 10,000 詞頻帶。

會這些詞是學習者能使用目標語的關鍵指標。5000 詞頻帶含有教師值得花更多時間教學生的詞彙，10,000 詞頻帶則是含有目標語的低頻詞彙。另外，VLT 的起初設計目標是為測量外籍大專生的學術英語能力，故亦採用學術詞彙表為測量範圍。

表 1：VLT 測試題目範例

1 business	
2 clock	<u>6</u> part of a house
3 horse	<u>3</u> animal with four legs
4 pencil	<u>4</u> something used for writing
5 shoe	
6 wall	

VLT 每個測試題皆含六個目標詞與三個釋義，並要求受測者從六個目標詞中選出符合釋義的三個詞（參見表 1）。每個部分包含 36 個目標詞及 18 個釋義（總共 6 個測試題目），均以隨機抽樣的方式選取（剔除專用名詞和複合詞），此測試結果較能評估受測者於各詞頻帶中所學會的詞彙比率有多高，進而推論其總詞彙量。

2.3.2 英語詞彙深度知識測試工具：Word Associates Test (WAT)

在詞彙深度知識研究方面，詞彙聯想測試（Word Associates Test，WAT）為英語常用的詞彙深度知識測試工具。與上述 VLT 不同的是 WAT 均以高頻詞彙為目標詞。此工具共有 50 個測試題，各含一個目標詞及八個選項詞。此外，每個測試題的選項詞被分成兩組，各四個，一組為聚合關係¹¹詞彙的選項，另一組則是組合關係¹²詞彙之選項（見表 2）。WAT 每題最高分為 4 分，表示受測者成功選出四個正確的聯想詞。雖然各測試題均有四個正確答案，但不一定由兩個聚合關係詞彙與兩個組合關係詞彙組成，即有可能由一個聚合關係詞彙和三個組合關係詞彙所組成，反之亦然。

¹¹ 聚合關係（Paradigmatic Relationship）是指目標詞的同義詞或近義詞，或與目標詞的意思有關之詞彙，如 “sudden” – “quick, surprising”。

¹² 組合關係（Syntagmatic Relationship）是指通常與目標詞共現的詞彙，亦即搭配詞，如 “sudden” – “change, noise”。

表 2：WAT 測試題目範例

sound							
logical	healthy	bold	solid	snow	temperature	sleep	dance

3. 研究方法與材料

3.1 華語語料庫詞頻統計調查

本文所用的詞頻統計為「現代漢語語料庫詞頻統計¹³」，其提供「中央研究院現代漢語平衡語料庫」的詞頻訊息，並收錄根據出現頻率排序之第 1 到 93,826 位的詞彙。然於設計測試工具前，得先根據該詞頻統計調查華語學習者之詞頻概貌（詳見 2.1 節「詞彙頻率」）。

華語學習者詞頻概貌的調查範圍與 Milton (2006) 及 Richards 等人 (2008) 的研究相同，包括屬於 5000 詞頻帶的詞彙，以亂數表隨機抽樣之方式，從每 1000 詞頻帶中各抽出 20 個詞彙為目標詞，並剔除專有名詞（如新加坡、國民黨、臺北等）與感嘆詞（如啊、喔、嘿等），最後共選出 100 個目標詞。此外，本研究採“Eurocentres 10K Vocabulary Size Test (EVST)”為調查方式 (Meara and Jones 1990, 引自 Read 2000)，讓學習者於詞彙單勾選其認識或學過之詞彙，以期透過短時間測試較大量的詞彙。然而，EVST 的缺點在於無法驗證學習者的作答是否具有自我高估的傾向，故除了 100 個目標詞外，同時加入假詞以作測試。整體分數之計算方式如下：（一）每個針對真詞的「是」，表示學習者知悉真詞之意義，得一分；（二）每個針對假詞的「是」，表示學習者認識假詞之意義，扣一分。

本研究所用的假詞共 35 個，佔目標詞數量的 35%，即 7 個假詞彙代表每 1000 詞頻帶。設計假詞的過程包括以下步驟：（一）從每個詞頻帶抽出一些詞彙作為設計假詞之參考（如從 1000 詞頻帶抽出「中國」與「之前」）；（二）從抽出的參考詞彙得知哪些單字或詞素曾出現於該詞頻帶，因此可用作製作假詞的成分（如「中」、「國」、「之」、「前」）；（三）使用不同參考詞彙的單字或詞素合成一個假詞（如將「中」與「前」組成假詞「中前」）。

詞頻概貌調查本研究採線上測試方式，要求學習者於問卷點選「是」或「否」，以代表其是否認識螢幕上的目標詞。調查對象不限國籍及母語背

¹³ <http://elearning.ling.sinica.edu.tw/CWordfreq.html>

景¹⁴，共計 96 位華語學習者，包括 12 位初級、16 位初中級、19 位中級、26 位中高級與 23 位高級學習者。調查結果顯示華語學習者詞頻概貌（圖 2）的確與 Meara（1992）所提出的語言學習者典型的詞頻概貌（詳見圖 1）有著相似趨勢，亦即各詞頻帶的柱狀從左到右呈現下斜傾向。因此，我們所採用「現代漢語語料庫詞頻統計」的初步調查結果確實可作為設計華語學習者詞彙廣度與深度知識測試工具的基礎。

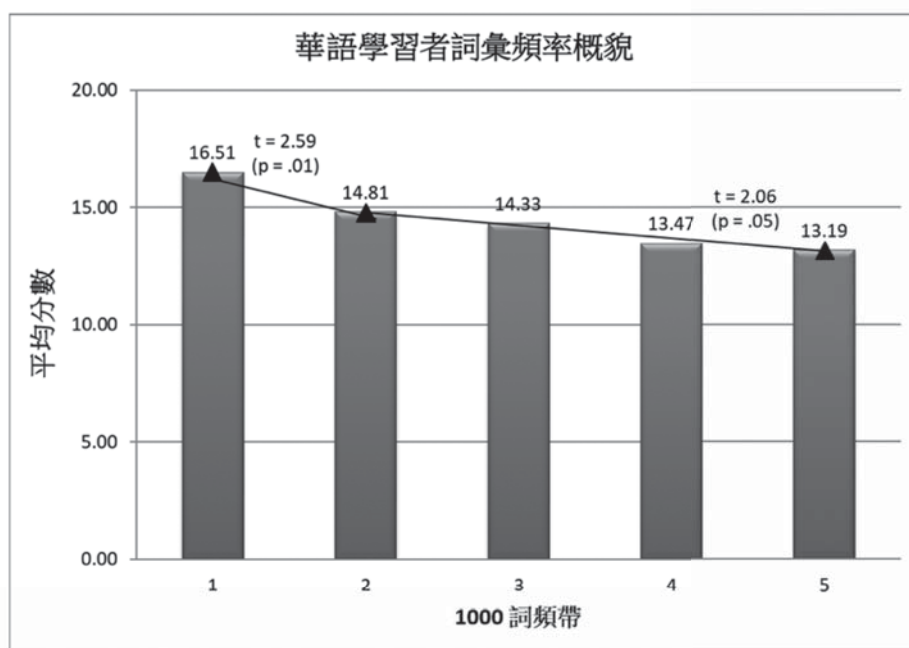


圖 2：華語學習者詞頻概貌¹⁵

¹⁴ 受測者的國別分布如下：亞洲國家（如馬來西亞、韓國等）76.04%；歐洲國家（如拉脫維亞、挪威等）8.33%；美洲國家（如美國、尼加拉瓜等）13.54%；非洲國家（如聖多美普林西比）2.08%。受測者母語分布如下：印尼語21.88%；泰語16.67%；日語和英語各13.54%；越南語9.38%；韓語和馬來語各5.21%；西班牙語4.17%；法語2.08%；葡萄牙語、俄語、土耳其語、拉脫維亞語、挪威語、緬甸語、荷蘭語和雙語（荷蘭語、西班牙語）各1.04%。

¹⁵ 有關本研究的詞頻概貌顯著性檢定（t 檢定）顯示，1000 與 2000 詞頻帶（ $t = 2.59$, $p = .01$ ）、2000 與 5000 詞頻帶（ $t = 2.06$, $p = .05$ ）皆達到顯著差異，表示華語學習者對 1000 詞頻帶的詞彙明顯較為熟悉，而對 5000 詞頻帶詞彙的熟悉度顯著較低。

另一方面，在設計測試工具前，本文亦根據詞彙覆蓋率將「現代漢語語料庫詞頻統計」分成五個等級，並與 VLT 採用的英國國家語料庫（British National Corpus, BNC）之詞頻等級作一比較，如下表 3 所示。跟 BNC 相同，本文將「現代漢語語料庫詞頻統計」第 1 和第 2 的 1,000 個詞彙歸類為第 1 個等級，只是累積的覆蓋率（68.08%）比 BNC（86.06%）稍低。基本上，本文進行等級分類時，須確保下一個等級的覆蓋率不比上一個等級的覆蓋率高，如此才可較合理地呈現屬於下一個等級詞彙的出現頻率比上一個等級詞彙的出現頻率來得低。根據以上原則，表 3「現代漢語語料庫詞頻統計」每個等級的覆蓋率如下：（一）第 1 等級覆蓋率 68.08%；（二）第 2 等級覆蓋率 7.92%；（三）第 3 等級覆蓋率 4.23%；（四）第 4 等級覆蓋率 1.47%；（五）第 5 等級覆蓋率 1.32%。此詞頻等級將成為設計華語學習者詞彙廣度與深度知識測試工具的基礎¹⁶。

表 3：「現代漢語語料庫詞頻統計」與 BNC 詞頻等級比較

詞頻排序	「現代漢語語料庫詞頻統計」			BNC		
	覆蓋率	累積覆蓋率	等級	覆蓋率 (%)	累積覆蓋率	等級
	(%)	(%)	¹⁷	(%)	(%)	
第 1 的 1,000 個詞	59.38	59.38		77.96	77.96	
第 2 的 1,000 個詞	8.70	68.08	1	8.10	86.06	1
第 3 的 1,000 個詞	4.72	72.80		4.36	90.42	2
第 4 的 1,000 個詞	3.20	76.00	2	1.77	92.19	
第 5 的 1,000 個詞	2.35	78.35		1.04	93.23	3
第 6 的 1,000 個詞	1.88	80.24	3	0.67	93.90	
第 7 的 1,000 個詞	1.47	81.70	4	0.45	94.35	
第 8 的 1,000 個詞	1.32	83.02	5	0.33	94.68	4

¹⁶ 本文的設計範圍僅包括第 1 和第 2 等級，期望未來後續研究能延續探究第 3 級到第 5 等級，並設計更完整的華語學習者詞彙廣度與深度知識測試工具。

¹⁷ 與 VLT 不同，本文的測試工具不包括學術詞彙表，每個部分的目標詞都來自於「現代漢語語料庫詞頻統計」，因此詞頻統計也與 BNC 不同，不僅分成四個等級，而是分成五個等級。

3.2 華語學習者詞彙廣度知識測試工具之設計考量

如 2.2 節「詞彙知識」所述，本文所界定的詞彙廣度知識是指 Meara (1996) 的詞彙量組成。按照 Nation (1990) 的詞彙組成，詞彙量組成包含「書寫形式」¹⁸及「概念義」兩個基本組成。因此，在測試形式方面，本文採取 VLT 為設計基準，著重測量目標詞與釋義的搭配，亦即受測者除需知道詞語的書寫形式外，更須了解該詞的概念義，此與本文詞彙量組成的定義相符。另外，受測者若要正確作答，不但要了解目標詞之義，亦要了解每個釋義選項的意思，故能以少量題目測出大量詞彙，適合用作推論和估計詞彙量。

上表 3 整理 8,000 個詞彙於不同詞頻帶的覆蓋率與等級，然考量測試時間，測試工具的設計範圍僅包括第 1 和第 2 等級，即語料庫中出現頻率排序為第 1 至 4,000 位的詞彙。目標詞的選擇採取亂數表隨機抽樣，並從每 2000 詞頻帶抽出 150 個詞，包括 60 個名詞（40%）、78 個動態與狀態動詞（52%）、12 個副詞（8%）¹⁹。由於每個測試題含六個目標詞，故每 2000 詞頻帶的測試題目將有 10 個名詞題目、13 個動詞題目、2 個副詞題目，共 25 題。

統計上，以 2,000 個詞彙為母群體，若設定為 95% 信賴水準（ $p = .05$ ）與 5% 抽樣誤差，我們必須從每 2000 詞頻帶中選出至少 333 個詞彙樣本，才能較準確地推論或估計詞彙量。然而，設計包含 333 個目標詞的測試工具，不但數量過多，不甚實用；而且受測者為了成功答對每個測試題答，不但要了解每個目標詞之義，更需了解釋義所使用的每個詞彙的意思。有鑒於此，本研究主張把 333 個詞彙樣本運用於測試工具的目標詞及釋義中，如此即可隨機以 25 個測試題估計學習者在每 2,000 詞彙範圍的詞彙量。

¹⁸ 因考量表達性詞彙知識測試的複雜度，本研究測試範圍僅含接受性詞彙知識之測試，即詞形方面僅包括書寫形式，不包括口說形式。此外，因考量華語字詞是表義文字，為二語學習者的學習難點之一，因此測試題同時呈現漢字形式及漢語拼音，如此才能避免測試工具僅能測出學習者的認字能力，而無法達到原本測量詞彙知識所設定的目標。

¹⁹ 本文設定測試題目僅包含名詞、動詞、副詞，因為根據華語八千詞詞類分佈情況，此三個詞性累積的比率已高達 92.38%。此外，三個詞性在測試工具中的比率亦根據其於華語八千詞詞類分佈的比率（名詞 42.06%、動詞 51.24%、副詞 6.70%）來決定（張莉萍 2007）。

設計釋義的參考資料有二，一是教育部《重編國語辭典修訂本》²⁰，二是由中文詞彙網路（Chinese Wordnet, CWN）語料庫製作的《中文字典》²¹。設計原則有四：（一）釋義須為目標詞最基本的意思，故要避免使用含有比喻或任何延伸義的釋義；（二）目標詞及釋義所使用的詞彙必須屬於同一詞頻帶，且釋義盡量使用詞頻比目標詞高的詞彙，如此才能減低受測者知道目標詞但不了解釋義之可能性；（三）釋義不能出現目標詞裡面所含有的任何字，尤其能夠提供作答線索的字詞；（四）釋義盡量以簡單的語法形式呈現，以免受測者受到太多閱讀干擾的影響²²。

表 4：華語學習者詞彙廣度知識測試工具之試題範例

1 問 wèn	
2 去 qù	② 往、離開 wǎng, líkāi
3 聽 tīng	① 向人請教 xiàng rén qǐngjiào
4 帶 dài	⑤ 做好、結束事情 zuò hǎo, jiéshù shìqíng
5 完成 wánchéng	
6 選擇 xuǎnzé	

綜上所述，本研究所用之華語學習者詞彙廣度知識的測試工具共 50 題，一半代表 2000 詞頻帶的測試題，另一半則是代表 4000 詞頻帶之題目。在每個測試題中，受測者必須選出三個符合釋義的目標詞（範例見上表 4），因此每題最高分數為 3 分，滿分是 150 分。所得分數代表受測者於華語 4,000 詞彙範圍的詞彙量比率，如 150 分等於 4,000 詞彙量、75 分等於 2,000 詞彙量，如此類推。

²⁰ 資料來源：<http://dict.revised.moe.edu.tw/cbdic/>。

²¹ 資料來源：<http://chinese.cdnet.info/>。

²² 本研究盡量使用《實用視聽華語》第一冊語法點來設計釋義，僅有小部分釋義以《實用視聽華語》第二冊語法點設計，如「把」字句：「S + 把 + O + V 來」。

3.3 華語學習者詞彙深度知識測試工具之設計考量

Meara (1996) 所指的第二個詞彙知識組成，亦即本文所界定之詞彙深度知識，是詞彙組織組成，除包含 Nation (1990) 的「聯想詞」組成（或 Richards (1976) 的「詞彙網絡中的聯想詞」組成）之外，更涉及 Nation (1990) 的「搭配方式」組成。故於設計測試工具時，本研究以英語詞彙聯想詞與搭配詞測試方面最具有影響力的 WAT 為測試形式。

深度知識測試工具的設計範圍比廣度知識小，僅限於 2,000 個頻率最高的詞彙，即就是屬於第 1 等級之詞彙。此設定除考量測試時間外，亦考慮到語言學習者實際習得詞彙之情況。李曉 (2007) 發現詞彙廣度與深度知識雖然呈高度正相關，但詞彙深度知識的發展仍落後於詞彙廣度知識，此意味著學習者需較長時間累積深度知識，故測量深度知識時，應把重點放在高頻詞彙，如此才能得出更精確的結果。

雖然許多英語詞彙廣度與深度知識的研究以 VLT 與 WAT 為測試工具，但由於兩種測試工具是由學者們透過不同研究過程設計而成，因此二者測試結果的關聯性仍較為模糊，原因在於（一）兩種工具的目標詞不同，無法直接比較學習者對某些詞彙的廣度與深度知識；（二）僅觀察 WAT 的測試結果，研究者無法判斷學習者的錯誤在於不了解目標詞與聯想詞的關聯（缺乏深度知識），抑或尚未掌握目標詞義（缺乏廣度知識）所致。為了設計一套較具整體性的測試工具，本研究的詞彙深度測試工具亦使用詞彙廣度測試中的目標詞。除能使兩種測試結果之關聯性較為直接外，詞彙深度的測試結果亦不會受學習者對目標詞的掌握程度影響，研究者可以先從詞彙廣度的測試結果觀察學習者是否掌握目標詞。

於 2000 詞頻帶方面，詞彙廣度測試工具的目標詞共有 150 個，但真正與釋義搭配僅 75 個，含 30 個名詞、39 個動態與狀態動詞及 6 個副詞，然 Piper 與 Leicester (1980) 的研究結果指出學習者與母語者在名詞聯想詞任務表現較為一致，並無顯著差異，意味著名詞聯想詞對學習者而言較易掌握（引自 Wolter 2001）。因此，我們認為詞彙深度測試工具的內容應該與動詞與副詞為主，即 39 個動詞與 6 個副詞，不含任何名詞²³。

²³ 本研究於詞彙深度知識測試中省略名詞內容，主要原因在於試題數量與測試時間之考量。

本研究設計之詞彙深度知識所測試的聯想詞與目標詞具有以下三種關係：

（一）聚合關係（**Paradigmatic Relationship**），即聯想詞是目標詞的同義詞或近義詞，如「漂亮」與「美麗」，或與目標詞某方面意思有所關聯，如「大」與「老」（年紀大也是「老」的意思）；（二）組合關係（**Syntagmatic Relationship**），即聯想詞為目標詞的高頻共現搭配詞，如「完成」與「任務」；（三）解析關係（**Analytic Relationship**），即聯想詞代表目標詞的某方面或成分，亦有可能是目標詞詞典意思的一部分，如「廣告」與「吸引」（「廣告」於《中文字典》之釋義為「將商品用文字、圖畫宣告於大眾，或登載於報章雜誌，或經由電視、收音機的傳播，或印成傳單小冊散發，以吸引顧客，推銷貨物」）。

本文以教育部《重編國語辭典修訂本》及 CWN 的《中文字典》作為選擇聚合關係與解析關係聯想詞之參考，組合關係聯想詞則以《中央研究院現代漢語平衡語料庫》²⁴ 為選擇範圍。此外，選項詞（包括聯想詞及其它錯誤選項）之選取遵循以下主要原則：（一）選項詞與目標詞必須屬於同一詞頻帶，且盡量使用詞頻比目標詞高的詞彙，始能降低學習者認識目標詞但不知道選項詞而錯誤作答之可能性；（二）為免提供過於明顯的作答線索，聯想詞不可含有目標詞成分，如以「製作」為「製造」的聯想詞；（三）詞彙深度測試所用的聯想詞並未出現於同一目標詞之詞彙廣度測試釋義中。

詞彙深度知識測試工具共計 45 題，各測試題共有四個正確答案，最高分數為 4 分（範例見下表 5）。此外，每個受測者可能得到的最高分取決於受測者在詞彙廣度知識測試的表現。當廣度測試結果顯示受測者尚未掌握某個目標詞時，該目標詞將不被納入深度測試的分數計算範圍之內，故深度測試的結果不會受到詞彙量不足的影響。詞彙深度知識的得分以百分比計算，結果反映於受測者的詞彙量中，有百分之幾的詞彙在詞彙網絡建立了正確的組織，分數計算範例如表 6。

表 5：華語學習者詞彙深度知識測試工具之試題範例

去 qù							
<input type="checkbox"/> 高 gāo	<input checked="" type="checkbox"/> 走 zǒu	<input type="checkbox"/> 長 cháng	<input checked="" type="checkbox"/> 死 sǐ	<input type="checkbox"/> 積極 jījí	<input type="checkbox"/> 不足 bùzú	<input checked="" type="checkbox"/> 玩 wán	<input checked="" type="checkbox"/> 旅遊 lǚyóu

²⁴ 資料來源：<http://asbc.iis.sinica.edu.tw/>。

表 6：詞彙深度測試分數計算範例

受測者	詞彙廣度測試	詞彙深度測試		
	正確答案	最高得分	實際得分	百分比
A	45	180	90	50%
B	20	80	80	100%

上表 6 顯示 A 學習者在廣度測試表現非常好，答對 45 個目標詞，表示其已掌握深度測試中 45 個目標詞的基本詞義。由於深度測試每題（每個目標詞）最高分數是 4 分，故 A 學習者在深度測試可能得到的最高分為 180 分。然而，A 的實際得分僅 90 分，因此其詞彙深度知識百分比為 50%，表示在 A 的詞彙量中，僅有 50% 詞彙在詞彙網絡中建構正確的組織。另一方面，B 學習者雖於廣度測試表現較差，但因為在深度測試成功拿到最高得分，故其詞彙深度知識百分比是 100%，反映他對詞與詞之間的關係（如近義關係、搭配關係）相當了解，亦即具有較為結構性的詞彙知識。

4. 主要研究結果與討論

4.1 華語學習者詞彙廣度與深度知識測試工具之信度與效度

本研究採取折半信度（Split-half Reliability）以及史布公式（Spearman-Brown Prophecy Formula）以考察測試工具的信度。由於本測試工具的題目皆根據難度順序排列²⁵，故於計算折半信度時，會將測試題分成奇數與偶數各半²⁶並考察二者的相關係數。其後，所得的折半信度（奇數與偶數試題的相關係數）將使用史布公式調整或轉換成整體測試信度，兩種測試工具的信度詳見表 7。

表 7：詞彙廣度、深度知識測試工具之信度

	折半信度	整體信度
詞彙廣度知識	0.98	0.99
詞彙深度知識	0.97	0.99

²⁵ 測試題的難度排序依照目標詞的詞頻而定，高頻目標詞的難度比低頻目標詞低，反之亦然。

²⁶ 當測試題根據難度順序排列時，將測試題分成奇數與偶數兩半來計算折半信度是最為理想的處理方法（王文科與王智弘 2014）。

表 8：詞彙廣度、深度知識測試工具的因素分析結果

	隱性變項	總方差解釋
詞彙廣度知識	1. 名詞詞彙量 2. 動詞、副詞詞彙量	72.39%
詞彙深度知識	1. 聚合關係、組合關係聯想詞知識 2. 解析關係聯想詞知識	63.09%

效度考察方面，本文採取因素分析（Factor Analysis）²⁷ 調查測試題究竟蘊含幾個因素或隱性變項（Latent Variables）。按照測試工具設計的建構定義，因素分析的結果應該只能從每個測試工具提取單一因素，如「詞彙量」為廣度測試的因素以及「聯想詞知識」為深度測試的因素²⁸，此為最理想的情況，即代表該測試工具確實達到其測量目標，反映測試結果能單純估計學習者的詞彙量與聯想詞知識，並無受到其它語言能力的干擾。然而，因素分析的結果顯示本文兩種測試工具各含至少兩個²⁹ 被測量的因素或隱性變項。進一步以因素負荷量（Factor Loading）³⁰ 分析每個測試題，得悉每個隱性變項與詞彙量或聯想詞知識存在密切關聯，且總方差解釋（Total Variance Explained）皆高於 60%，反映本研究的兩種測試工具仍具有一定效度（參見表 8）。

²⁷ 本文已檢驗符合因素分析的兩個基本假設：（一）KMO 取樣適切性量數（KMO Measure of Sampling Adequacy）大於 0.60，表示樣本的共同因素足以進行因素分析；（二）Bartlett 球形檢定（Bartlett's Test of Sphericity）結果顯示，樣本屬於多變量常態分配（Multivariate Normal Distribution），因此可以進行因素分析。

²⁸ 因素分析提取而來的因素或隱性變項到底為何，並無統一的解讀或明確答案（Schmitt et al. 2001），因此研究者僅能依測試工具設計的建構定義提出可能的假設。

²⁹ 因素分析陡坡圖（Scree Plot）的陡坡部分是從第二個因素開始。

³⁰ 因素負荷量（Factor Loading）是指每個測試題與各因素的相關係數。

4.2 華語學習者詞彙廣度與深度知識的測試結果

本研究詞彙廣度與深度知識測試的受測者為 53 位華語學習者，分別來自臺灣各大學、或國語中心的華語學習者或學過華語的外籍學位生（不限國別與母語³¹）。測試結果顯示，本文詞彙廣度與深度知識的測試工具可有系統地區分四個不同程度的華語學習者，包括初級、初中級、中級、中高級學習者³²。即在 95% 信賴區間的情況下，各程度學習者之間的表現皆達顯著差異（詳見表 9）。至於中高級與高級學習者之表現無顯著差異，推測其原因在於廣度知識測試工具以 4,000 個詞彙的設計範圍，無法涵蓋中高級以上學習者之詞彙量，也就無法客觀地卻分中高級受試者。若按照 TOCFL 與 HSK 的詞彙量標準，中高級學習者的詞彙量已能達到 5,000 個詞彙（參見附註 5），因此，本研究 4,000 個詞彙的設計範圍自然難以估計中高級以上的詞彙量。

表 9：華語學習者詞彙廣度知識測試的平均分數、估計詞彙量及顯著性檢定

程度	樣本數	平均分數 (滿分 = 150)			估計詞彙量 ³³			顯著性檢定
		2000 詞頻帶	4000 詞頻帶	總數	2000 詞頻帶	4000 詞頻帶	總數	
初級	11	21.91	15.82	37.73	584	422	1006	
初中級	7	45.29	37.29	82.57	1208	994	2202	$t = 3.33 (p = .05)$
中級	14	62.36	59.71	122.07	1663	1592	3255	$t = 3.06 (p = .05)$
中高級	13	70.77	70.46	141.23	1887	1879	3766	$t = 2.75 (p = .05)$
高級	8	70.13	71.50	141.63	1870	1907	3777	$t = 0.11$ (無顯著)

³¹ 受測者的國別分布如下：亞洲國家（如馬來西亞、韓國等）77.36%；歐洲國家（如英國、義大利等）11.32%；美洲國家（如美國、尼加拉瓜等）9.43%；非洲國家（如聖多美普林西比）1.89%。受測者母語分布如下：越南語 35.85%；日語 15.09%；英語 11.32%；蒙古語 9.43%；韓語 5.66%；西班牙語、泰語和印尼語各 3.77%；葡萄牙語、俄語、斯威士語、義大利語、法語和馬來語各 1.89%。

³² 受測者的華語程度根據以下標準分成初級、初中級、中級、中高級、高級五個等級：（一）TOCFL 成績；（二）HSK 成績；（三）學習的時間長短；（四）學習的環境（CFL 或 CSL 環境）；（五）學習者自我評價。

³³ 估計詞彙量 = (平均分數 / 75) × 2000，如：初級 2000 詞頻帶的估計詞彙量 = (21.91 / 75) × 2000 = 584.27 ≈ 584；初級 4000 詞頻帶的估計詞彙量 = (15.82 / 75) × 2000 = 421.87 ≈ 422。

由於本文的詞彙廣度知識測試工具以至少 333 個詞彙作為 2,000 個詞彙的樣本，因此估計詞彙量在統計上已達 95% 的信賴水準（參閱第 4 節）。此外，詞彙廣度知識的測試結果（表 9）顯示初級、初中級和中級學習者的估計詞彙量確實符合 TOCFL 與 HSK 詞彙量標準³⁴：（一）初級估計詞彙量為 1,006，對應 TOCFL 入門級 500~1,000 個詞彙及 HSK 三級 600~1,200 個詞彙；（二）初中級估計詞彙量為 2,202，對應 TOCFL 基礎級 1,000~2,500 個詞彙及 HSK 四級 1,200~2,500 個詞彙；（三）中級估計詞彙量為 3,255，對應 TOCFL 進階級與 HSK 五級 2,500~5,000 個詞彙。

表 10：華語學習者詞彙深度知識的測試結果及顯著性檢定

程度	樣本數	2000 詞頻帶 的 詞彙量 ³⁵	詞彙深度知識的測試結果			顯著性檢定
			平均分數 (滿分=180)	詞彙深度 知識 ³⁶	估計 2000 詞頻 帶具有深度知識 的詞彙量	
初級	11	584	25.82	52.19%	305	
初中級	7	1208	74.71	69.92%	845	$t = 3.64 (p = .05)$
中級	14	1663	114.04	81.81%	1361	$t = 2.86 (p = .05)$
中高級	13	1887	157.69	93.72%	1768	$t = 3.32 (p = .05)$
高級	8	1870	161.25	96.13%	1799	$t = 0.19$ (無顯著)

在詞彙深度測試方面，有一個值得注意的現象是，測試結果可能會受到學習者詞彙量的影響，即學習者因詞彙量不足而出現深度知識被低估的情況。簡單來說，當學習者在詞彙深度測試的表現不佳時，問題未必出於其詞彙深

³⁴ TOCFL 與 HSK 的詞彙量標準（資料來源：<http://www.sc-top.org.tw/chinese/LS/test5.php>）是提供每個考試等級所需要的最低詞彙量為參考，因此每個等級的詞彙量範圍是該等級的詞彙量至該等級以上的詞彙量，譬如參加初中級（TOCFL 基礎級）考試的學習者必須具備至少 1,000 個詞彙量，參加中級（TOCFL 進階級）考試的學習者則必須具備至少 2,500 個詞彙量，那麼初中級學習者的詞彙量範圍是 1,000 至 2,500 個詞彙。

³⁵ 由於深度知識測試工具的設計範圍僅限於 2,000 個高頻詞彙，因此估計具有詞彙深度知識的詞彙量也僅按照 2000 詞頻帶的估計詞彙量，詳如表 9 所示。

³⁶ 詞彙深度知識的百分比取決於學習者在詞彙廣度知識測試的表現，詳細的計算方法請參閱第 5 節「華語學習者詞彙深度知識測試工具的設計」。

度知識之不足，而是缺乏詞彙量所致。換言之，當學習者作答錯誤時，不見得是他不知道目標詞的聯想詞或搭配詞，而是他根本不認識該目標詞之義。若無法解決此問題，詞彙深度知識測試工具恐怕只是測試詞彙量或詞彙廣度知識的間接工具，亦即未能真正評估出學習者的詞彙深度知識，僅能代表詞彙量之多寡（高分數代表詞彙量多、低分數代表詞彙量少）。

為免詞彙深度知識的測試結果受到詞彙量所影響，本研究於設計測試工具與進行測試兩方面，遵循以下三個原則：（一）測試工具的目標詞亦是詞彙廣度知識測試工具的目標詞；（二）參加詞彙深度測試前，受測者必須參加詞彙廣度測試；（三）當受測者在詞彙廣度測試選錯某個目標詞的答案，該目標詞不再包含於詞彙深度測試分數的計算範圍。此外，學習者的詞彙深度知識以百分比呈現，以表示於學習者的詞彙量中，有百分之幾的詞彙具有深度知識，例如：在中級學習者的 1,663 個詞彙量當中，有 81.81%（1,361 個）的詞彙具有深度知識（參見表 10）。由此可見，詞彙廣度與深度知識測試工具之整合設計使得測試結果更具描述廣度與深度知識關聯性之功能。測試結果亦顯示：華語學習者之詞彙廣度與深度知識之間，存在高度正相關，且相關係數（Pearson r ）達 0.952，高於英語詞彙廣度與深度知識的中度相關性（Schmitt and Meara 1997；Nurweni and Read 1999；李曉 2007）。本文推測此差異的原因在於測試工具的整合設計。此現象於 Qian（2002）、Qian 與 Schedl（2004）的研究可見一斑，當兩種測試工具使用同一目標詞表時，測試結果間的相關性存在偏高傾向。Qian（2002）考察“Test of English as Foreign Language-Vocabulary Measure（TOEFL-VOC）”與 WAT 的相關性，發現兩種測試結果的相關係數為 0.68。然當 Qian 與 Schedl（2004）將 WAT 的目標詞改成 TOEFL-VOC 的目標詞時，測試結果間之相關性則高達 0.84。

另外，線性迴歸分析（Linear Regression Analysis）³⁷ 結果說明詞彙廣度知識（自變項）與詞彙深度知識（依變項）的迴歸方程式具有顯著性（ $F(1,44) = 422.832, p < 0.01$ ），而詞彙廣度知識對於詞彙深度知識具有 90.6% 的預測力（ $R^2 = 0.906$ ）。華語學習者詞彙深度知識的預測值為 $37.047 + 0.015$ 詞彙廣度知識，亦即於 0~4,000 個詞彙量的範圍內，每增加 1,000 個詞彙量，詞

³⁷ 本文透過離群值診斷（Outliers Diagnostics）剔除 7 個離群值。剩下 46 個受測者的資料符合線性迴歸分析的基本假設：（一）線性關係（Linear Relationship）；殘差獨立性（Independence）或無自我相關（No Auto-Correlation）；（三）殘差同方差性（Homoscedasticity）；（四）殘差常態性（Normality）。

彙深度知識則增長 15%。

4.3 華語學習者詞彙廣度與深度知識測試工具的修訂

除探討詞彙廣度與深度知識的關係之外，本文也根據一些不理想的測試結果分析及檢測試題設計。根據詞彙廣度知識的測試結果（見表 9），可以發現高級學習者 4000 詞頻帶的平均分數竟然高於 2000 詞頻帶（然兩者無顯著差異， $t = 0.83$ ）。我們進一步分析發現：在 8 位高級受測者當中，其中 4000 詞頻帶平均分數高於 2000 詞頻帶者共有 6 位，而只有在第 13 個測試題的某個項目，受測者答錯比率高達 66.6%（4 位）。該測試題包含一個釋義項目「不用再等」及六個目標詞為答案的選擇「不必」、「總是」、「至少」、「立刻」、「的確」、「另外」，此測試題的正確答案為「立刻」，4 位答錯的受測者皆選擇「不必」。由此可見，「不必」為可能誤導受測者的選項，此或與其含義和釋義的「不用」相近所致。

除第 13 題外，6 位受測者在另外兩個測試題的表現亦較不理想，分別是第 10 題的釋義項目：「以為確實如此，不懷疑」及第 25 題的釋義項目：「動作或事件發生迅速」。第 10 題的正確答案是「相信」，而 3 位（50%）答錯的受測者中，2 位選擇「實際」，另外 1 位則選擇「敢」，顯然「實際」為誤導受測者的主要選項，其產生誤導情況之原因與「不必」相同，因為釋義中含有與錯誤答案「實際」意思相近的「確實」。第 25 題正確答案為「忽然」，而 3 位（50%）答錯的受測者中，2 位選擇「即將」，1 位則選擇「畢竟」，可見「即將」為誤導的主要選項。據 CWN《中文字典》，「即將」之釋義為「事件會在很短的時間內發生」，其與「忽然」的釋義有至少兩個相同的成分——「事件」、「發生」，且「在很短的時間內」亦可能被解讀為「迅速」。由此可見，「忽然」的釋義雖無直接出現與錯誤答案「即將」的意思相近的字詞，但仍包含三個與其詞典釋義成分相同或有關之成分，故容易造成誤導。然而，因「即將」亦是第 25 題的目標詞之一，因此修改「即將」等於另外再選一個目標詞。此外，我們發現第 17 題其中一個釋義項目「不同的地方」，有 2 位（33.3%）受測者皆錯誤選擇「區域」，而非正確答案「差異」，推測原因可能出於釋義內容包含與錯誤答案「區域」意思相近的字詞——「地方」。

在釋義設計方面，我們已盡量避免使用出現於目標詞的成分，即可能提供正確答案線索的字詞。然而，上述發現為測試工具之設計帶來重要啟示——在撰寫詞彙廣度知識測試工具之釋義時，必須注意釋義內容應避免可能產生誤導作用的成分，包括錯誤選項所含的單字、意思與錯誤選項相近的字詞、

或與錯誤選項釋義成分相同或有關之成分。簡言之，測試工具不僅要避免給予有利作答的線索，亦應盡量減少錯誤選項所產生的誤導作用³⁸。有鑒於此，本文針對以上 4 個設計不良的測試題提出修訂及建議，詳列如表 11。完整的華語學習者詞彙廣度與深度知識測試工具請參閱雷達莉（2016）。

表 11：華語學習者詞彙廣度知識測試題目之修訂

測試題號	修訂項目	原先設計	修訂建議
10	錯誤選項	「實際」	「值得」
13	錯誤選項	「不必」	「原來」
17	錯誤選項	「區域」	「模式」
25	目標詞 釋義	「即將」 「會在很短的時間內發生」	「顯然」 「觀察到的事件很明確」

整體而言，上述之研究成果仍有些許侷限與無法全面涵蓋的面向，希望未來其他研究仍可朝以下三點建議改善。在受測者人數方面，本研究僅有 53 位受測者參與，仍屬於小規模研究，雖然沒有違反統計假設，但難免會有較大抽樣誤差的風險；爾後研究盼能有更多受測者人數參與，以降低抽樣誤差的風險外，更可再確認測試工具信度與效度，檢視結果。另一方面，本研究之測試工具的設計範圍僅限於 4000 詞頻帶以下的詞彙，導致測試結果無法涵蓋中高級以上學習者之詞彙量。因此，建議後續研究可以設計代表 4000 詞頻帶以上詞彙的詞彙知識測試工具（詞頻等級請參見表 3），以使研究者能更完善了解各程度華語學習者詞彙知識的情況。最後，在詞彙知識與語言能力之關係研究，本文希望能以標準的華語詞彙知識測試工具來促進華語詞彙知識與語言能力關係之相關研究，如探討詞彙廣度與深度知識分別與各項語言技能（聽力、口說、寫作、閱讀）之相關性的研究。

5. 結語

本文以英語詞彙知識測試工具為參考基準，設計兩種華語學習者詞彙知識的測試工具，包括：（一）詞彙廣度知識測試工具，參考英語 Vocabulary Level Test（VLT），以估計（95% 信賴水準）華語學習者的詞彙量；（二）

³⁸ 然而，假如幾位被誤導的受測者（僅指選擇「不必」、「實際」、「即將」或「區域」四個選項的受測者），均選擇正確答案，高級 4000 詞頻帶的平均分數（71.50）仍高於 2000 詞頻帶的平均分數（71.38），只是兩者的差別變小。因此，有關高級受測者平均分數的問題亦與隨機誤差有部分關聯。

詞彙廣度知識測試工具，以英語 Word Associates Test (WAT) 為參照對象，藉此了解華語學習者的詞彙組織。通過信度考察結果顯示，本文所設計之測試工具顯然具有高信度，不論是詞彙廣度抑或詞彙深度的測試工具，信度均高達 0.99。而在效度方面，因素分析從兩個測試工具各提取出兩個被測量的隱性變項，此兩個變項於廣度和深度的總方差解釋分別為 72.39% 與 63.09%，可算具有一定效度。

根據初步測試結果，兩種測試工具均能區分四個程度華語學習者的詞彙知識，包括初級、初中級、中級以及中高級以上程度之學習者（共 53 位受測者）。中高級與高級學習者的表現無顯著差異，原因或與中高級與高級學習者的詞彙量本身已超出測試工具設計範圍的上限（4,000 個詞彙）有關。除統計上已存在 95% 信賴水準外，初級、初中級和中級學習者（不包括詞彙量可能已超出 4,000 個詞彙的中高級及高級學習者）的估計詞彙量均符合 TOCFL 或 HSK 之詞彙量標準，證明詞彙廣度知識測試工具的精準性。

此外，兩種工具之整合設計使得測試結果間具有更直接的關係，可用來描述詞彙廣度與深度知識的關聯性。從相關性考察的結果可發現本文的詞彙廣度與深度知識存在高度正相關（ $r = 0.952$ ），且廣度知識對深度知識具有 90.6% 預測力。此結果意味著提升詞彙深度知識必須從習得大量新詞彙著手，亦顯示詞彙的重要性，乃為語言能力的重要核心成分。

引用文獻

- Barrow, Jack. 1999. Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey. *System* 27.2: 223-247.
- Laufer, Batia, and Zahava Goldstein. 2004. Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning* 54: 399-436.
- Meara, Paul. 1992. *EFL Vocabulary Tests*. University College Swansea: Centre for Applied Language Studies.
- Meara, Paul. 1996. The dimensions of lexical competence. *Performance and Competence in Second Language Acquisition*, eds. by Gillian Brown, Kirsten Malmkjaer, and John Williams, 35-53. Cambridge: Cambridge University Press.
- Milton, James. 2006. X-Lex: The Swansea vocabulary level test. *Proceedings of the 7th and 8th Current Trends in English Language Testing (CTELT) Conference Vol. 4*, eds. by Christine Coombe, Peter Davidson, and Dwight Lloyd, 29-39. UAE: TESOL Arabia.

- Milton, James. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Nation, Paul I.S. 1990. *Teaching and Learning Vocabulary*. New York: Heinle and Heinle.
- Nation, Paul I.S. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nurweni, Ari, and James Read. 1999. The English vocabulary knowledge of Indonesian university students. *English for Specific Purpose* 18.2: 161-175.
- Pignot-Shahov, Virginie. 2012. Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers* 4: 37-45.
- Qian, David D. 2000. Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review* 56.2: 282-308.
- Qian, David D. 2002. Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning* 52.3: 513-536.
- Qian, David D., and Mary Schedl. 2004. Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing* 21.1: 28-52.
- Read, James. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Richards, Brian, David Malvern, and Suzanne Graham. 2008. Word frequency and trends in the development of French vocabulary in lower-intermediate students during Year 12 in English schools. *Language Learning Journal* 36.2: 199-214.
- Richards, Jack C. 1976. The role of vocabulary teaching. *TESOL Quarterly* 10: 77-89.
- Schmitt, Norbert, and Paul Meara. 1997. Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition* 19: 17-36.
- Schmitt, Norbert, Diane Schmitt, and Caroline Clapham. 2001. Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing* 18.1: 55-88.
- Wolter, Brent. 2001. Comparing the L1 and L2 mental lexicon a depth of individual word knowledge model. *Studies in Second Language Acquisition* 23: 41-69.

- 王申。2010。〈詞彙廣度與深度對英語專業學生閱讀的影響〉，《雲南農業大學學報》，第4卷第2期，59-62。[Wang, Shen. 2010. The difference between the effects of vocabulary depth and those of vocabulary size on reading comprehension. *Journal of Yunnan Agricultural University* 4.2: 59-62.]
- 王文科、王智弘。2014。《教育研究法》。臺北：五南。[Wang, Wen-Ke, and Chi-Hong Wang. 2014. *Educational Research Method*. Taipei: Wunan.]
- 白麗梅、陳艷玲。2005。〈英語詞彙知識的廣度和深度在閱讀理解中的作用〉，《河西學院學報》，第21卷第4期，53-56。[Bai, Li-Mei, and Yan-Ling Chen. 2005. The roles of breadth and depth of vocabulary knowledge in reading comprehension. *Journal of Hexi University* 21.4: 53-56.]
- 李曉。2007。〈詞彙量、詞彙深度知識與語言綜合能力關係研究〉，《外語教學與研究》，第39卷第5期，352-359。[Li, Xiao. 2007. Assessing the roles of breadth and depth of vocabulary knowledge in second language proficiency. *Foreign Language Teaching and Research* 39.5: 352-359.]
- 張莉萍。2007。〈華語文能力測驗(TOP-Huayu)發展現狀〉，發表於「外語能力測驗之動向與展望國際學術研討會」，2007年7月。臺北。[Zhang, Li-Ping. 2007. Current development of test of Chinese as a foreign language (TOP-Huayu). Paper presented at *International Conference on "Trends and Prospects of Foreign Language Proficiency Testing"*, July 2007. Taipei.]
- 辜玉旻、陳咏馨、林千惠、張苑真。2012。《重新思考詞彙知識在國小低年級學童閱讀理解中的角色》。行政院國家科學委員會專題研究計畫成果報告(NSC100-2420-H-008-001-)。桃園：國立中央大學。[Gu, Yu-Min, Yong-Xin Chen, Qian-Hui Lin, and Wan-Zhen Zhang. 2012. *Rethinking the Role of Vocabulary Knowledge in the Reading Comprehension of Early Grade School Students*. Taoyuan: National Central University. (NSC100-2420-H-008-001-)]
- 雷達莉。2016。《華語學習者詞彙廣度與深度知識測試的設計》。國立臺灣師範大學碩士論文，臺北。[Mahastuti, Lestari. 2016. *Test Design of Breadth and Depth of Vocabulary Knowledge for Chinese Learners*. Taipei: National Taiwan Normal University MA thesis.]
- 錢旭菁。2002。〈詞彙量測試研究初探〉，《世界漢語教學》，第4卷，54-62。[Qian, Xu-Jing. 2002. A preliminary study of vocabulary size test. *Chinese Teaching in the World* 4: 54-62.]

華語文教學研究

[審查：2017.11.30 修改：2018.2.23 接受：2018.3.23]

雷達莉

Lestari MAHASTUTI

10610 臺北市大安區和平東路一段 162 號

國立臺灣師範大學華語文教學系暨研究所

Department of Chinese as a Second Language

National Taiwan Normal University

No. 162, Sec. 1, Heping E. Rd., Daan Dist., Taipei City 10610, Taiwan

lestarimahastuti@gmail.com

蕭惠貞

Hui-Chen S. HSIAO

10610 臺北市大安區和平東路一段 162 號

國立臺灣師範大學華語文教學系暨研究所

Department of Chinese as a Second Language

National Taiwan Normal University

No. 162, Sec. 1, Heping E. Rd., Daan Dist., Taipei City 10610, Taiwan

huichen.hsiao@ntnu.edu.tw

A Preliminary Test Design of Breadth and Depth of Vocabulary Knowledge for Chinese L2 Learners

Lestari MAHASTUTI

Hui-Chen S. HSIAO

**Department of Chinese as a Second Language
National Taiwan Normal University**

Abstract

Most previous studies on vocabulary knowledge have primarily focused on English learners' vocabulary knowledge, with established and standardized English vocabulary knowledge test instruments employed as main research tools (Qian 2000, 2002; Bai and Chen 2005; Li 2007; Wang 2010). Nevertheless, the lack of standardized vocabulary knowledge test in Mandarin field remains as one issue that lead to the falling behind of Chinese vocabulary knowledge research. In view of this, the present study aimed to fill the gap by designing a preliminary test of breadth and depth of vocabulary knowledge specifically for L2 Chinese learners, with reference to the existing English vocabulary knowledge test instruments (cf. VLT and WAT).

Overall, the present results showed the characteristics of newly designed instruments illustrate the following unique features: (1) High reliability (both breadth and depth instrument achieved high split-half reliability of 0.99); (2) Considerable degree of validity (using factor analysis, two latent variables were extracted from both breadth & depth instrument, with the total variance explained of 72.39% and 63.09% respectively); (3) Accurately (95% confidence level) estimate the vocabulary size; (4) Distinguish and Identify four different proficiency levels (basic to post-intermediate) of learners. Moreover, the integrated design of two test instruments allows the test results to be clearly interpreted in terms of the relationship between L2 learners' breadth and depth of vocabulary knowledge.

Keywords: vocabulary knowledge, breadth of vocabulary, depth of vocabulary, vocabulary test design, TCSL