

## 影響對外漢語閱讀測驗難度的因素

### 摘要

本文旨在探討對外漢語短文閱讀測驗中，哪些因素會影響試題的難度。研究人員以華語文閱讀能力測驗 (TOCFL) 為研究素材，在文本、提問兩大項目下，列出數個變項，並將這些變項跟每道試題的難度參數進行多元迴歸分析，以找出對試題難度具有預測力的變項。結果顯示，在閱讀文本中，中等詞彙數、副詞數、高等詞彙數、二重複句句長、以及三重複句 (含以上) 句長等五個變項對試題難度具有預測力，而提問方式對試題難度則無明顯影響。

關鍵字：對外漢語測驗、TOCFL、預測難度、難度分析、閱讀能力

### 1. 文獻探討：

一般探討第二外語閱讀測驗難度的文獻，以英文閱讀測驗居多。Freedle and Kostin (1993) 從提問 (test item)、文本 (test text)、提問及文本互動性 (text-by-item interaction) 三個方向來分析 TOEFL 的閱讀測驗，每一項下再細分數個變項，最後發現有七項因素明顯影響試題難度<sup>1</sup>。Alderson (2000) 在論及影響閱讀測驗難度的因素時，主要從影響提問 (test item) 的因素及影響文本 (test text) 的因素兩大方面著手，前者涵蓋了提問所使用的語言、提問類型、應試技巧、文法、詞彙在測驗中的角色、是否能使用辭典、以及讀者智力等因素；後者則涵蓋了讀者的背景知識、回答問題時所需閱讀的文本、文章長度及題組形式等等。之後，在論及歐洲線上語言診斷測驗 DIALANG 時，Alderson (2005) 做出更細緻的區分，從文本 (test text)、讀者 (reader)、提問/任務 (Questions/Tasks) 三方面來思考可能影響閱讀測驗難度的因素。徐惠琦 (2010) 則指出影響英文閱讀理解難度的三個因素，分別為閱讀材料、閱讀題型、以及其所考察的閱讀技能。

至於探討漢語作為外語的閱讀測驗難度的文獻則較少，部分研究只針對文本或試題類型做處理，或是研究的規模較小。張寧志 (2000) 曾就幾項漢語教材的語料做定量分析，針對不同難度的文本，約略列出每百字句數、平均句子長度、非常用詞數等指標。吳門吉 (2005) 在考察影響留學生漢語閱讀因素的研究中，以試題類型為主，列出難度由高而低的試題種類。王佶 (2006) 則以 HSK 基礎閱讀測驗為研究素材，從閱讀材料、閱讀題幹、閱讀選項三方面來探討影響閱讀難度的可能因素。此外，台灣師範大學的可讀性團隊開發了文本可讀性指標自動

---

<sup>1</sup> 此七項因素為語義重疊度、句長、句段長、段落長、文章組織、否定詞及指示詞的使用。

化分析系統<sup>2</sup> (2011), 內含詞彙、語意、句法、文章凝聚性四類自動化指標, 雖然尚未以對外漢語閱讀測驗為重心, 卻是一項能快速分析漢語文本可讀性的工具。

綜觀各學者的研究方向, 可看出在影響閱讀測驗難度的研究上, 多半區分為三個區塊: 文本內容、提問類型、讀者因素。在影響文本內容的因素上, 通常先從語言結構 (如詞彙頻率、句子長度、特殊句型、連接詞數、否定詞數等等) 著手, 之後才會擴展至語篇風格, 如文體、題材、作者風格、文化因素等項目。在提問類型的部分, 則著眼於不同類型的試題以及其所考察的閱讀技能, 如王佑 (2006) 將不同題幹所測試的閱讀能力分為三類: 捕捉主要事實和重要細節的能力、領略文章大意的能力、對作者態度、情緒或文章內容做出推斷的能力; 徐惠琦 (2010) 則將英文閱讀測驗所能測知的能力區分為 11 項。換句話說, 不同的試題類型可測試不同的閱讀技能, 這也可能成為決定閱讀測驗難度的因素。至於讀者因素, 取決於讀者方的語言能力、學習過程、知識背景等因素, 變項既多且雜, 目前較少有人專文研究, 本研究也暫時不處理這個部分。

本研究試圖找出影響對外漢語閱讀測驗難度的可能因素, 如過去的相關研究, 也區分成兩大方向: 一以「試題文本」為主, 列出各項可能影響文本的變項, 以統計方法探討各變項和試題難度的關係, 若有某變項對試題難度具預測力, 也意味這個變項能影響文本及試題難度。二以「試題提問」為主, 將各種提問歸類後, 觀察不同類型的題目是否明顯影響試題難度。本研究也將觀察各類型的提問在不同試題難度間的分布狀況、檢視不同程度的考生在答題過程中所表現的閱讀能力是否有明顯差異。本研究希望藉著上述幾項議題的探討, 達到下面目標:

- (1) 檢視目前華語文閱讀測驗 (TOCFL-Reading) 的題庫現況及應改進的方向。
- (2) 為各等級閱讀測驗的命題、審題人員提供更明確的命題建議, 以期適當掌握試題難度及測驗方向。

## 2. 研究方法:

### 2.1 研究素材:

本研究的素材取自華語文閱讀測驗 (TOCFL-Reading)<sup>3</sup> 經全國預試後的短文閱讀題目, 內含流利級 (C1) 文本 128 篇、子題數 300 題; 高階級 (B2) 文本 83 篇、子題數 157 題; 進階級 (B1) 文本 47 篇, 子題數 81 題。合計

<sup>2</sup> 文本可讀性指標自動化分析系統 1.0 的網址為 <http://www.chinesereadability.net/index.aspx>。

<sup>3</sup> 華語文閱讀測驗 (TOCFL-Reading) 按歐洲語言共同參考架構 (CEFR) 分為六個等級: 入門 (A1)、基礎 (A2)、進階 (B1)、高階 (B2)、流利 (C1)、精通 (C2)。

閱讀文本共 258 篇，子題數 538 題<sup>4</sup>，總字數約十萬字左右（含子題提問、選項），全部皆為多項選擇式的選擇題。在此有兩點須加以說明：

(1) 本研究主要的重心在短文閱讀，因此不包括像是廣告、公告、啟事、簡章等等的真實材料題。我們認為短文閱讀和材料閱讀在文本性質、體裁的差異大，閱讀目的及答題策略也不同，因此不宜放在一起比較。本研究各等級在文本篇數及子題數上的差異也由此而起，進階級(B1)的真實材料題居多，短文閱讀題較少；流利級(C1)則相反，真實材料題相當少，主要的試題內容皆為短文閱讀，因此其篇數及子題數最多。

(2) 這些題目雖然出自不同等級的預試卷，但因每一份預試卷皆置有共同題，因此不同等級的每道試題皆可透過等化的方式，在同一個難度量尺上做比較。不同等級的卷子經過統計串連之後，每道試題皆有一個試題難度參數，約在-2.5 至+3 之間，負值越小表示題目越簡單，正值越大表示題目越難。之後的研究步驟將以每道試題的難度參數為依變項，觀察其和不同分析項目(自變項)之間的關係。

## 2.2 研究步驟：

首先，我們從「閱讀文本」和「子題提問」兩個方向來設定分析項目：

**2.2.1. 閱讀文本的分析項目：**在閱讀文本方面，我們依照漢語詞彙、句法及語篇的特性，列出「詞彙構成」、「句法結構」、「銜接手段」三大分析項目，每一項目下各包含不同的變項：

A. **詞彙構成：**包含「初等詞彙數」、「中等詞彙數」、「高等詞彙數」等變項。

**說明：**詞彙難度是決定文本難易度的基礎因素之一。這裡的詞彙等級（初等、中等、高等）主要是依據華測會所公布的華語八千詞<sup>5</sup>而訂。每篇文本依斷詞系統斷詞後，分別計算各等級詞彙在該篇文本所佔的數量。

B. **句法結構：**包含「二重複句的句子長度」、「三重複句以上（含多重）的句子長度」等變項。

<sup>4</sup>在華語文閱讀測驗中，每一篇文本依其長度、難度的不同，可搭配一至五個子題。

<sup>5</sup>華語八千詞表請見華測會網站：<http://www.sc-top.org.tw/chinese/download.php>

**說明：**漢語的段落是由單句及複句所構成。在複句方面，又可依其結構形式和邏輯意義細分為一重複句、二重複句、三重或三重以上的多重複句。在本研究中，我們以人工方式分析各篇文本，標示出各文本中的二重複句或三重以上（含多重）的複句，並計算這些複句在該文本的總字數<sup>6</sup>。

**C. 銜接手段：**包含「連接性詞語數」、「副詞數」、「代名詞數」、「特殊零代詞數」等變項。

**說明：**儘管漢語語篇會採取替代、省略、連接、指稱銜接、名詞銜接等等銜接手段，但在本研究中，因需處理的文本數量及篇幅龐大，最後仍以「連接」及「指稱銜接」這兩種結構明顯、容易量化的銜接方式為主<sup>7</sup>。各篇文本經中研院斷詞系統斷詞之後，計算每篇文本的連接性詞語數<sup>8</sup>、副詞數<sup>9</sup>及代名詞數。「特殊零代詞數」此變項是由研究人員一一檢視文本後，由人工標記出該篇文本特殊零代詞用法<sup>10</sup>的次數。

**2.2.2. 子題提問的分析項目：**華語文閱讀測驗在研發及設計上，主要參考歐洲語言共同參考架構 (CEFR)，因此本研究在提問的分類上，我們也採用了該架構下的概念，將透過提問所測得的閱讀能力分為「了解主旨」(to identify main ideas)、「辨認細節」(to find specific details)、「做出推論」(to make

---

<sup>6</sup> 舉例來說，若某篇文本有兩個二重複句，而這兩個二重複句加起來的總字數為 60 字，則本篇文本的二重複句句長為 60 字。

<sup>7</sup> 「連接性詞語數」、「副詞數」屬「連接」方式；「代名詞數」、「特殊零代詞數」屬「指稱銜接」方式。

<sup>8</sup> 這裡的「連接性詞語」，主要是指能連接句子與句子，或是段落與段落的連接性詞語。依中研院平衡語料詞類的簡化標記（見 <http://ckipsvr.iis.sinica.edu.tw/> 的詞類標記列表），包括了 Cab（連接詞，如：等等）、Cba（連接詞，如：的話）、Cbb（關聯連接詞）、以及 Dk（句副詞）這幾類詞，但不包括 Caa（對等連接詞）。

<sup>9</sup> 我們知道在漢語中，並非所有副詞都起銜接上下文的作用，然而礙於文本數量及篇幅的龐大，無法一一精細區分，只能簡略以「副詞」類一以概之。這裡的「副詞」，是指文本經中研院斷詞系統斷詞後，依中研院平衡語料詞類的簡化標記，被標記為 D（副詞）的詞彙，但不包括 Da（數量副詞）、Dfa（動詞前程度副詞）以及 Dfb（動詞後程度副詞），因我們認為這幾類副詞基本上不起連接句子的功能。另外，由於 Dk（句副詞）一類在概念上已列入連接性詞語，我們並未將其列入副詞這一類。

<sup>10</sup> 根據張黛琪（2003）「零代詞的診斷式測驗與評量」一文，零代詞在句中充任的角色可以是主語、賓語或定語，但以充當主語的情況佔絕大多數（p.42），因此本研究僅觀察零代詞在文中充當主語的情況。該文將零代詞的結構做了多層次的分析（見該文表 3.1, p.44），除了單句之外，餘者都屬於篇章層次。我們的預設是，較複雜的零代詞使用方式，有可能影響文章難度，因此研究人員檢視文本，將其中以零代詞充當主語、但使用方式較特殊的部分標出（包括張黛琪一文所指出的承前跨句、蒙後一句、蒙後跨句以及多元現象這幾種用法），統計在一篇文本中特殊零代詞的出現次數。

inferences) 三大類<sup>11</sup>。在每一類能力之下，又依提問類型、提問和選項的關係、答題時主要使用的能力，進一步分出數個細項，見表一。研究人員以人工方式，將 538 道子題依表區分出 A (了解主旨)、B (辨認細節)、C (做出推論) 三類。

---

<sup>11</sup> 此三分類詳見 Dutch Grid: <http://www.lancs.ac.uk/fss/projects/grid/>、以及 Alderson(2005), p.125.

表一：提問及相應閱讀能力分類：

閱讀能力		可能提問	
<b>A</b> 了解主旨/資訊/目的 To understand/identify main idea	1 找主旨	本文/本段主要在說什麼？	
	2 下標題 (含部分衡量能力)	哪一個適合作為本文的標題？	
	3 辨認作者 (或文中人物) (已言明的) 立場/看法	作者的立場是什麼？	
	4 了解完整的故事情節	這個故事的主角怎麼了.....？	
<b>B</b> 找出特定細節或資訊 To find specific details or information	非針對性的細節	1 肯定問法	下面哪一個是對的？
		2 否定問法	下面哪一個是錯的？/...不包括什麼？
	針對性的細節	3 找特定方法或原因	如何...？為什麼...？
		4 找特定情況/行動/反應/問題...etc. 或解釋) 特定詞語/短語	在什麼情況下？最擔心什麼？..... 什麼叫做「...」？
<b>C</b> 根據文章內容做出推論 To make inferences	1 找出 (未明說的) 條件、前提、觀察角度、結論....etc.		在什麼前提下，作者認為...？ ...從哪一個角度來看這件事...？
	2 詮釋詞或短語或短句 (名詞替代的用法)		文中所提到的「這一點」是指什麼？ (或將文中某句劃線，重新解釋)
	3 依據文本找出解決方法		根據本文，我們應該...怎麼做？
	在文中做歸納比較	4 肯定問法	關於...，現在和過去有什麼不同？ ...跟...這兩種理論的共同點是什麼？
		5 否定問法	根據本文，比較各個方法，哪一個不適合做為...？
	6 推論作者 (或文中人物) 的意圖、寫作態度		作者為什麼提到....？
	7 推論文本未來走向		這篇文章的最後，如果要以一句話作結束，下面哪一句話最適合？
	8 歸納全文組織結構		關於本文的寫作方式，哪個是對的？

	9 根據故事情節（未言明的部分）做推論	故事中的主角為什麼....？
--	---------------------	----------------

我們將上述所列出的分析項目整理成表二。如表二的例子所示，每一道子題有一個難度參數，和這道子題搭配的閱讀文本則有分析項目 1-10，而這道子題也依其提問性質，可歸類到 A、B、C 三種能力中的其中一種。我們以這些分析項目為預測變項，以每道子題的難度參數為效標變項，進行逐步迴歸分析 (stepwise multiple regression) <sup>12</sup>。結果如表三所示。

表二：納入逐步迴歸的分析項目

			例一	例二	
<b>該子題的難度參數 (效標變項)</b>			<b>1.234</b>	<b>1.539</b>	
<b>納入逐步迴歸的分析項目 (預測變項)</b>	<b>閱讀文本的分析項目</b>	1. 該文本總字數		316	623
		詞彙構成	2. 初等詞彙數	59	124
			3. 中等詞彙數	50	77
			4. 高等詞彙數	10	27
		句法結構	5. 二重複句句長	240	261
			6. 三重複句 (含以上) 句長	0	0
		銜接手段	7. 連接性詞語數	10	7
			8. 副詞數	34	34
			9. 代名詞數	1	10
			10. 特殊零代詞數	0	2
	<b>子題提問的分析項目</b>	主要能力 A (了解主旨)		●	
		主要能力 B (辨認細節)			●
		主要能力 C (做出推論)			

<sup>12</sup> 多元迴歸分析用於根據好幾個自變項預測一個依變項的情形，目的在了解哪些自變項對於依變項較有預測力。在自變項的輸入方法上，又可分為逐步迴歸 (stepwise)、順向選擇 (forward selection)、反向剔除 (backward elimination) 和強迫輸入法 (enter) 四種，逐步迴歸法為實務上較常使用的方法。



表三<sup>13</sup>：分析項目對試題難度預測的逐步迴歸分析摘要表

次序	投入變項	R	R <sup>2</sup>	$\Delta R^2$	淨 F 值	迴歸係數	標準化迴歸係數	t 值
1	中等詞彙數	0.622	0.387	0.386	338.185***	.027	.600	8.153***
2	副詞數 <sup>14</sup>	0.636	0.404	0.402	15.445***	-.023	-.281	-4.768***
3	高等詞彙數	0.645	0.416	0.413	10.791**	.020	.138	2.654**
4	二重複句句長	0.650	0.423	0.419	6.493*	.002	.122	2.923**
5	三重複句(含以上)句長	0.654	0.428	0.422	4.393*	.001	.104	2.096*

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

### 3. 研究結果及討論：

按研究結果(表三)，對試題難度具有預測力的分析項目，依序有「中等詞彙數、副詞數、高等詞彙數、二重複句句長、以及三重複句(含以上)句長」五個項目被選入。其中(見 $\Delta R^2$ 一欄)，中等詞彙數可預測 38.6%的試題難度；若加上副詞數，預測率可達 40.2%的，加上高等詞彙數，可達 41.3%；加上二重複句句長，可達 41.9%；加上三重複句(含以上)句長，五個變項的整體預測率則達 42.2%。這五個變項對閱讀文本的難度具預測力，亦可視為能影響閱讀文本難度的因素。

我們可從以下兩點來討論這個結果：

<sup>13</sup>投入變項部分，原結果尚有第六個變項「總字數」被選入，但因六個投入變項模式中，「總字數」的容忍度(tolerance)數據偏低，數值為 0.082，變異數膨脹因素(variance inflation factor, 簡稱 VIF)為 12.187，條件指標(condition index, 簡稱 CI)為 20.824。容忍度數值介於 0 至 1 之間，若容忍度太小，表示此變項和其他自變項間有共線性(collinearity)問題；變異數膨脹因素是容忍度的倒數，若值越大，越有共線性問題；條件指標為如果在 15 以上，則表示可能有共線性問題，如果在 30 以上，則表示有嚴重共線性問題(吳明隆, 2000)。上述結果皆顯示「總字數」此一變項與其他被選入之變項之間可能有共線性存在，也就是與其他變項相關太高，會導致迴歸分析的估計不準確。積差相關分析結果發現，此變項與 Intermediate 和副詞數的相關係數分別為 0.909 ( $p < 0.001$ ) 與 0.908 ( $p < 0.001$ )，故研究者最後刪除「總字數」變項，選擇投入五個變項之預測模式。

<sup>14</sup>雖然中等詞彙數與副詞數的積差相關也偏高，為 0.806 ( $p < 0.001$ )，然而條件指標為 10.636，小於 15，且研究者認為此二變項本質上仍為不同之概念，可能只是由於中等詞彙內，包含相當數量的副詞，才導致變項間的關聯性較強。

3.1. 顯然詞彙構成仍是決定一篇閱讀文本難度的主要因素。但值得注意的是，中等詞彙對文本難度的決定性大於高等詞彙。據我們觀察，這可能是因為多數功能詞大多出現在中等詞彙中，學習難度較高，而高等詞彙多半是低詞頻的實詞，不具語法功能，因此對試題難度的決定性略低。從試題編寫的角度來看，過去我們通常認為在較難的文本中，高等詞彙的比重應較大，但這個結果顯示，即使一篇文本絕大多數由中等詞彙組成，仍可能是一篇相當難的文本。

在銜接手段方面，則只有副詞一項被選入變項，但這一項在解釋上卻較為困難。當初受限於斷詞系統的分類，我們只排除了不具句子連接功能的數量副詞 (Da)、動詞前程度副詞 (Dfa)、動詞後程度副詞 (Dfb)，以及已納入連接性詞語的句副詞 (Dk)，其他全納入副詞一類<sup>15</sup>。但並非所有納入副詞的詞語都具連接功能，因此此項研究結果無法斷定具連接功能的副詞越多，文章難度就越高。

當初我們抱著嘗試的心態列出副詞一項，沒想到這項卻被選入影響文章難度的變項，這個結果也許能從以下兩點來說明：一、從統計上來看，副詞這個變項和中等詞彙的相關度高，因此兩者可能有牽引關係，意即中等詞彙多、副詞數也增加，副詞一項因而被選入，這也再度說明了中等詞彙所包含的功能詞（如具句法功能的副詞）較多，對文章難度的影響較大。二、漢語的副詞系統相當複雜，張誼生（2000）曾將副詞區分為兩大塊：「以詞彙義為主」和「以功能義為主」。若從前者的角度思考，一篇文章以詞彙義為主的副詞越多，表示修飾性高、意義豐富，的確有可能增加文章難度；若從後者的角度思考，一篇文章中具功能性的副詞越多，可能顯示語法較為複雜，也可能增加文本難度。由於本次研究的文本數量龐大，無法一一區分副詞的細類，因此僅能對統計結果做出以上推論，無法斷定是哪一類副詞可能影響文章難度。

在句法結構方面，則預設的兩項變項都被選入，顯示一篇文本若包含越多二重、或是三重以上的複句，對文本的難度影響越大。

3.2. 根據研究結果，一篇閱讀文本無論在詞彙構成、句法結構或是銜接手段上都有變項被選入預測模式，但是提問類型的變項（意即該題所測試的主要閱讀能力）對試題難度的影響卻沒被選入。換句話說，試題所測試的閱讀能力（若分為「了解主旨」、「辨認細節」、「做出推論」三大能力，見表一）和試題難度之間並無明顯關聯，這和華語文閱讀能力測驗原先預設的命題原則有所出入。我們原先設定，閱讀時的推論能力是較高級的技能、難度可能較高，因此在命題時，進階級（B1）沒有所謂的推論題，到了高階（B2）、流利級（C1）才逐步增加（見表四）。然

<sup>15</sup> 按中研院平衡語料庫的詞類標記，這裡的副詞包含了 Dab, Dbaa, Dbab, Dbb, Dbc, Dc, Dd, Dg, Dh, Dj 等幾類。

而，根據表四，儘管高階（B2）、流利級（C1）都有推論題，其難度平均值和其他兩種能力並沒有明顯的差異，可見推論題未必比較難。再看表五，即使以試題難度來區分，仍有不少推論題落在難度較低（如難度參數 0.68 以下）的區間，可見即使是閱讀能力較低的考生也能處理推論題。這個結果和歐洲線上語言診斷測驗 DIALANG 的結果一致（Alderson, 2005, p.137），它指出這三種預設的閱讀能力在 CEFR 各等級都很重要，而非如 CEFR 所設定，語言能力越高，推論能力就優於了解主旨或是辨認細節等能力。這主要的原因在於考生都是成年人，儘管語言能力有差異，但都具備一定的認知能力，因此無法從語言能力的高低來區分不同的閱讀能力。

表四：

子題所屬的等級	該等級的總題數	提問的分析項目	該分析項目的題數	該分析項目的題數佔該等級總題數的百分比	平均難度參數
流利（C1）	300	A（了解主旨）	57	19%	1.133
		B（辨認細節）	179	60%	1.017
		C（做出推論）	64	21%	1.421
高階（B2）	157	A（了解主旨）	39	25%	0.649
		B（辨認細節）	89	57%	0.906
		C（做出推論）	29	18%	1.002
進階（B1）	81	A（了解主旨）	21	30%	-0.914
		B（辨認細節）	60	70%	-0.85
		C（做出推論）	0		

表五：

子題的難度參數範圍	該難度範圍的總題數	提問的分析項目	該分析項目的題數	該分析項目的題數佔該等難度題數的百分比	平均難度參數
1.9-3 或 3 以上	71	A（了解主旨）	11	15%	2.357
		B（辨認細節）	44	62%	2.215
		C（做出推論）	16	23%	2.362
0.68-1.89 <sup>16</sup>	245	A（了解主旨）	51	21%	1.235
		B（辨認細節）	138	56%	1.24
		C（做出推論）	56	23%	1.421
0.68 以下	222	A（了解主旨）	55	25%	-0.331
		B（辨認細節）	146	66%	-0.324
		C（做出推論）	21	9%	0.126

<sup>16</sup> 這裡的 0.68，是預估通過高階級（B2）的考生能力值，1.89 則是預估通過流利級（C1）的考生能力值，因此暫以這兩個數值來劃分各等級的試題難度。

#### 4. 對華語文閱讀測驗的建議：

(1) 整體來看，詞彙構成及句法結構都各有兩項變項被選入，而銜接手段則只有副詞一項，而且副詞這一個變項還可能受到中等詞彙量的影響。換句話說，詞彙構成加上句法結構，似乎比量化的銜接手段（如連詞、副詞、代名詞的多寡）更能決定試題的難度。這項結果也許導因於兩種可能：一、漢語是意合的語言，比起形合的英語，對形式上的銜接手段依賴較少。二、漢語的銜接手段多元，但本研究僅僅就連接、指稱銜接等方式進行統計，未包括其他不易量化的替代、省略、名詞銜接等方式，因此有可能低估了其他銜接方式對於試題難度的影響。無論如何，以目前的研究結果來看，設定各等級的詞彙範圍仍是控制難度的重要手段。其中，以中等詞彙（約 3500 詞左右）的影響最大、高等詞彙（約 3000 詞左右）次之。此外，試題中是否出現多重複句、多重複句佔整體文本的多寡，也會明顯影響試題難度。然而，由於漢語的句限不明、銜接方式不明顯，命題者如不特意區辨，經常會忽略單複句的層次，對文章難度失去控制。

(2) 從表三、表四所呈現的數據來看，在華語文閱讀能力測驗中，辨認細節的題目過多，無論哪一個等級，這類題目幾乎都是過半的多數。我們認為這是一種不均衡的現象，如果如上（2.2.2）所述，了解主旨、辨認細節、做出推論這三種能力不應由語言程度來區分，在各等級試題中都很重要，那麼，在命題方向上，也應在各等級試題中，均衡地測試這三種閱讀能力才是。當然，進階級（B1）的文本由於內容淺顯、篇幅較短，產出推論型題目的機會較少，但不至於完全排除；至於高階（B2）、流利級（C1）的試題，多半是題組型態，一篇文本可附帶 3-6 個子題，那麼，在子題的命題上，就要考慮如何均衡分配（如哪幾題是考了解主旨、哪幾題是考辨認細節、哪幾題是考推論能力），才能測試到不同的閱讀能力。

(3) 本次研究雖納入近 260 篇的文本，然而大多數文體<sup>17</sup>皆為說明類，描述、記敘、論證類等文體的數量不多，而且完全排除指示類的文體（如廣告、公告等真實材料題），因此沒有將文體類型納入分析項目的變項中。在漢語測試中，文體不同是否會影響試題難度、不同等級試題所採用的文體及其比重是否應有區別，皆須進一步加以研究<sup>18</sup>。此外，漢語的銜接手段經常受到文體的影響，如零代詞

<sup>17</sup> 本研究在文體上採 CEFR 的分類，共五類：Descriptive（描述類）、Narrative（敘述類）、Expository（說明類）、Argumentative（論證類）、Instructive（指示類）。詳見 Dutch Grid：<http://www.lancs.ac.uk/fss/projects/grid/>。

<sup>18</sup> 文體類別和試題難度的關係，在〈On Hierarchies of Reading Skills and Text Types〉（1988）一文曾有相關研究，根據 ACTFL guidelines，不同語言等級原有其預設的適測文體。但研究發現，不同體裁的文本，其困難度和 ACTFL 預設的等級並不完全符合。在漢語測驗及 CEFR 的語言能力分級中，相關的研究仍付之闕如。

的使用通常受到篇章結構和文體的制約(廖秋忠、許餘龍)、連接性詞語在不同文體中的使用頻率也有所差異(劉志芳)。在本次研究中,銜接手段和試題難度之間除了副詞之外,並無明顯關聯。之後的研究,若能先區分文體,再比較其中的銜接手段與文本難易度,可能更有系統。這情形也指出華語文閱讀能力測驗在命題時須努力的方向,首先,應力求不同類型的文體在數量上達到均衡,才有材料來研究這些文體在不同等級的測試狀況,並觀察其對試題難度的影響力。

## 5. 結論：

本文對華語文閱讀能力測驗的命題工作提出幾點建議：在取材上，應注意各類文體在篇數上的均衡；在閱讀文本的編寫上，則應注意漢語文本的特性，適當控制文本難度；在編寫子題時，應考慮如何測試出不同閱讀能力並求取平衡。其中，在控制試題難度上，應從閱讀文本著手。根據研究結果，可能影響文本難度的因素有五項：中等詞彙數、副詞數、高等詞彙數、二重複句句長、以及三重複句(含以上)句長。了解這些因素對文本的影響，可讓華語文閱讀能力測驗在組織試題、控制難度、培訓命題教師時有所參考。至於提問部分，雖未明顯影響試題難度，卻能反映出該試題所欲測試的閱讀能力，也是命題教師、試題編審者所需注意的重點。

最後，本次研究限於研究材料、人力、時間，有幾點不足之處。在此簡要說明，以做為未來研究改進之參考：

(1) 如前所述，進階級(B1)的篇數較少，流利級(C2)篇數最多，各等級的篇數並不均衡。雖然每篇文本、試題都以共同題的方式串在同一個難度量尺上，我們仍擔心各等級篇數的不均衡有可能影響研究結果。這方面須等未來進階級(B1)的試題增加後，再持續納入研究。

(2) 影響閱讀文本難度的因素很多，本研究限於時間、資源、研究素材，無法充分列出各項變項。如台灣師大可讀型研究團隊所建置的「文本可讀性指標自動化分析系統」，在詞彙、句法、語意、文章凝聚性等指標下都列出多項變項。這些變項，哪些適用於在對外漢語閱讀的文本、哪些變項會對文本難度產生影響，都是未來可以持續研究的方向。

(3) 在本研究中，在閱讀能力的分類上僅區分為三類(即為前述「了解主旨」、「辨認細節」、「做出推論」三項)。在這三類下，可再區分出更細緻的項目(見表一)。理論上，每一個區分出來的細項(如閱讀能力A-1、A-2、B-3...等等)，都有助於了解考生的答題過程及相應的閱讀能力，也可觀察其和試題難度之間的關係。可惜，就目前題庫裏的題目來看，每道細項所具備的試題數差異甚大(比

發表於 2013 年加州中文教師學會(CLTAC) Spring Conference, Stanford University, March 9, 2013

如說閱讀能力 B-1 題數很多, C-7 題數則僅有一兩題), 還不足以做進一步研究, 須得未來題庫內題型更為豐富而多元之後, 才能進行相關研究。

### 參考書目

- 王佶 (2006) HSK (基礎) 閱讀理解難度的影響因素研究, 雲南師範大學學報 (對外漢語教學與研究版), 2006 年 5 月第 4 卷第 3 期。
- 徐惠琦 (2010) 影響閱讀理解難度的三個因素, 中小學外語教學 (中學篇) 2010 年第 1 期。
- 廖秋忠 (1992) 廖秋忠文集, 北京: 北京語言學院。
- 許餘龍 (2002) 對比語言學, 上海市: 上海外語教育出版社。
- 劉志芳 (2000) 連接性詞語於五種中文書面文體之使用。台灣師範大學華語文教學研究所碩士論文。 <http://ir.lib.ntnu.edu.tw/ir/handle/309250000Q/17137>。
- 張誼生 (2000) 現代漢語副詞研究。上海市: 學林出版社。
- 張黛琪 (2003) 零代詞的診斷式測驗與評量。台灣師範大學華語文教學研究所碩士論文。 <http://ir.lib.ntnu.edu.tw/ir/handle/309250000Q/17192>
- 張寧志 (2000) 漢語教材語料難度的定量分析, 世界漢語教學, 2000 年第三期。
- 吳門吉 (2005) 影響留學生閱讀因素的考察, 對外漢語閱讀研究, 北京大學出版社。
- 宋曜廷、張道行、陳茹玲、李宜憲 (2011): 可讀性特徵擷取自動化工具之發展。華語文自動處理與工具應用工作坊, 台北市: 國立臺灣師範大學, 2011 年 12 月 30 日。 <http://www.chinesereadability.net/index.aspx>。
- 吳明隆 (2003) SPSS 統計應用實務。台北: 松岡。
- Alderson, J.C. (2000). *Assessing Reading*. New York: Cambridge University Press.
- Alderson, J.C. (2005). *Diagnosing Foreign Language Proficiency*. London: Continuum.
- Alderson, J.C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR Construct Project. *Language Assessment Quarterly*, 3 (1), 3-30.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Council of Europe. (2009). *Relating language examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) .A Manual*. Strasbourg: Language Policy Division.
- Freedle, R., and Kostin, I. (1993) The prediction of TOEFL reading item difficulty: implications for construct validity. *Language Testing* 10, 133-170.
- Lee, J.F., and Musumeci, D. (1988) On Hierarchies of Reading Skills and Text Types. *The Modern language Journal*, Vol.72, No.2, p.173-187.