

## 嶺南大學「香港中文書面語水平測試」試卷質量研究 ——兼論對大學必修中國語文課程教學的啟示

李東輝 李斐 馬毛朋<sup>1</sup>

### 摘要

本文對嶺南大學自 2010 年以來的「香港中文書面語水平測試」試卷的質量進行了研究，數據分析結果顯示，該試卷的質量是良好的。如何讓檢測語言知識的題目也盡可能地反映出考生的語言運用能力，是未來試卷修改時需要關注的問題。本研究還發現，考生對普通話中的多音多義字和文言文的掌握最不理想，日後的教學需要加強有關方面的訓練；考生的文言文閱讀理解能力不足，未來在課程內容增加文言文，值得積極考慮；考生的現代文閱讀能力良好，教學中宜以訓練高階的閱讀能力即鑒賞、評價、創造能力為目標；考生具有中等或偏上的寫作水平，加強閱讀訓練不失為進一步提升寫作能力的方法。

**關鍵詞：**中文書面語 語言測試 大學中國語文教學

<sup>1</sup> 李東輝，嶺南大學校長室、中國語文教學與測試中心、研究生院。聯絡電郵：dhui@ln.edu.hk（本文通訊作者）。

李斐，嶺南大學中國語文教學與測試中心、陝西師範大學語言科學研究所。聯絡電郵：lifei@ln.edu.hk。

馬毛朋，嶺南大學中國語文教學與測試中心、陝西師範大學語言科學研究所。聯絡電郵：mp.ma@ln.edu.hk。

本研究獲得嶺南大學教學發展基金（項目號：102476）的資助。國家教育部語言文字應用研究所張一清研究員、富麗副研究員、陳菲助理研究員、北京師範大學朱寧寧博士在不同部分也參與了該研究項目，不僅對本研究提供了不少寶貴意見，也在數據分析方面提供了必要的幫助。兩名匿名審稿人的重要修改建議讓本文減少了疏漏。謹向以上諸位，致以誠摯的謝意！

## 一、研究背景

早在 1998 年，嶺南大學中文語言教學同仁就構想設立書面中文能力方面的「評核試」，用劃分等級的形式檢測畢業生的中文水平，以供未來的僱主參考（李雄溪，1998）。該考試的初期研製計劃於 1998 年獲得嶺南教學基金的支持，由胡維堯、李東輝負責完成並於 1998-1999 學年進行了第一次測試（胡維堯，2001）。嶺南大學中國語文教學與測試中心於 2000 年 11 月成立，中心創辦主任李東輝首次為嶺大引進教育部國家級普通話水平測試（PSC），而該中心也成為國家級普通話水平測試在香港的第五個考試點。至此，嶺大學生的「中文離校試」口試就是 PSC，而「中文離校試」書面語的試卷則由嶺南大學中心教師與教育部語言文字應用研究所（2009 年之前為「國家語委普通話培訓測試中心」）合作擬出，嶺大學生在畢業前根據個人需要自願參加。在十多年雙方合作編製的「中文離校試」試卷經驗的基礎上，2009 年，合作雙方決定繼續深化「香港中文書面語水平測試」的研究工作。國家教育部語言文字應用研究所姚喜雙所長和嶺大中文中心李東輝主任分別牽頭，合作開啟了由 29 人參與的中文書面語研究項目，是項研究也獲得了語委劉照雄教授和香港田小琳教授的大力支持與參與。該研究的目標是研製適合香港社會的中文書面語水平測試以及編纂以提升港人中文書面語能力為目標的學習綱要。前者作為衡量個人中文書面語能力的參考，將在嶺南大學學生中試行，未來希望也能夠推向社會。

該研究參考了嶺大書面語考試的情況、內地中文高考的試卷、香港教育局為中學設定的教學綱要和學習範疇等。其成果集中體現為：(1)「香港中文書面語水平測試」的評卷標準、題型研製以及根據、等級劃分描述以及樣卷；(2) 60 萬字的《香港中文書面語水平測試實施綱要》，該《綱要》包括以《通用規範漢字表》一級字為基礎的「字表」、以《普通話水平測試用詞語表》、《現代漢語常用詞表》和《教師語文能力評核（普通話）參照使用普通話詞語表》為基礎的「詞語表」、100 篇古詩文、140 篇現代文、語法修辭綱要、閱讀理解輔導及

寫作指南等內容。《綱要》涵蓋了書面語考試考核項目的大部分內容，之後的書面語試卷的出題工作，皆以此為據。2010 年 7 月召開了該學術研究成果的課題專家鑑定會，與會的七位專家對我們的研究成果給予了充分的肯定<sup>2</sup>，並訂立出該研究第二階段的研究方向。第一次考試也於 2010 年在嶺大進行。

從 2010 年至 2020 年，應試者均為嶺南大學學士學位課程的學生。2019 年，李東輝、李斐、馬毛朋獲得嶺南大學教與學研究基金的支持，作為該研究項目的第二階段，開始了「嶺南大學『香港中文書面語水平測試』試卷質量研究」。國家教育部語言文字應用研究所張一清、富麗、陳菲、北京師範大學朱寧寧都在不同程度上參與了該研究項目。自 2010 年積累的試卷和過千份考生答卷是用於本次研究的材料，希望本研究可以作為未來「香港中文書面語水平測試」進一步修改、完善的依據。

## 二、研究目標

本研究首要的目標是試卷的質量。評鑒試卷質量的常用指標包括信度（reliability）、效度（validity）、難度（difficulty index）、區分度（discrimination index）等。前兩項主要用來衡量試卷的整體質量。就效度來說，我們尚沒有用不同的辦法檢測應試者書面中文能力的有關數據。因此，這裏主要是結合試卷結構，討論內容效度（content validity）問題。每道題目的質量則體現為難度和區分度。我們知道，以上這些指標一般用於衡量標準化考試（standardized testing），一般不用於「水平測試」這種「標準參照測試」（criterion-referenced testing），衡量後者的一個指標是「可靠度」（dependability），即評卷者判斷應試者是否達到某個水平的內部一致性，可用卡帕係數（kappa coefficient）計算（Fulcher, 2010, p. 81）。由於計算卡帕係數需要同一批學生至少參加同一個測試兩次，我們尚無相關數據，而且更主要的

2 七位專家為陳章太、傅永和、張世平、于根元、葛蘭、王鐵琨、周建設。

是，我們希望對試卷各題目本身的情況有所了解，所以本研究仍然藉助對上述指標的分析，形成關於試卷質量的一些結論。

目前香港各專上學院，一般都有學士學位學生必修的中文課程，比如香港中文大學的課程名為「大學中文」，香港大學的課程名為「實用中文」，嶺南大學的課程為「中國語文傳意」，未來擬更名為「大學中文」，本研究通過對試卷的分析，也同時了解考生書面中文能力的狀況，希望對香港各大學必修中文課程的教學提出一些參考意見。

### 三、語言能力與試卷設計及內容效度

喬姆斯基 (Chomsky) 區分了語言能力 (competence) 和語言運用 (performance)。其對語言能力的定義常被人們認為不能解釋語言是如何使用的。海姆斯 (Hymes) 因此提出了交際能力 (communicative competence) 這一概念。後來喬姆斯基本人也引入了語用能力 (pragmatic competence) 的術語，即關於語言是如何在特定的語境中使用的知識 (Cook & Newson, 1996, p. 23)。語言能力既包括對一種語言的知識，也包括運用這種語言的能力，已是語言學家們的共識。Bachman (1990, pp. 81-89) 結合語言學與語言測試領域的研究成果，提出了著名的交際語言能力 (communicative language ability CLA) 的概念。其後，Bachman 又在論著中對有關概念進行詳述及微調，其基本框架則是一致的 (Bachman, 1991; Bachman & Palmer, 2010, pp. 44-52)。根據 Bachman, CLA 中的語言能力 (language competence) 分為組織能力 (organizational competence) 和語用能力 (pragmatic competence) 兩方面。前者又細分為語法能力 (grammatical competence, 指語音、詞彙、句法知識及組詞成句的能力) 和篇章能力 (textual competence, 指組句成篇的能力); 後者分為語義能力 (propositional competence, 決定話語如何表達命題內容)、功能能力 (functional competence, 決定話語如何表達語言使用者意圖)、社會語言學能力 (sociolinguistic competence, 決定話語如何與語言使用的場景特徵聯繫)。Bachman 對語言能力的這一分解，對語言測試研究與

教學來說，相當有實用性。我們在設計書面語試卷時，也是希望通過不同的題型，較為全面的檢測考生的書面中文能力。現將試卷各題與 Bachman 各項能力的對應列表如下：

表 1：題型與語言能力對照

能力題型	語言	語法能力	篇章能力	語義能力	功能能力	社會語言學能力
第一部分：用字辨析	1.1 選擇同音字	✓				
	1.2 給多音多義字注音	✓				
	1.3 辨別錯別字	✓				
第二部分：詞語辨析	2.1 完成常用成語	✓				
	2.2 選詞填空	✓	✓	✓	✓	✓
第三部分：辨別病句	辨別病句	✓	✓	✓	✓	✓
第四部分：閱讀理解	4.1 閱讀文言文	✓	✓	✓	✓	✓
	4.2 閱讀現代文	✓	✓	✓	✓	✓
第五部分：寫作	命題作文 / 實用文二選一	✓	✓	✓	✓	✓

從上表可見，Bachman 列舉的各項語言能力，在試卷中都有相應的題目加以檢測。

試卷的內容看，與香港中學文憑試中國語文試卷相比，本試卷包含檢測語音、詞彙、語法知識的題型。比如據《2021 年香港中學文憑考試評核大綱 (修訂版)》，文憑試「中國語文」公開考試卷一為「閱讀能力」，卷二為「寫作能力」，卷三為「聆聽及綜合能力考核」(其中「聆聽」部分考察的是口語理解能力，「綜合能力考核」部分考察的也是寫作能力); 據最新的《2024 年香港中學文憑考試評核大綱》，考試只有「卷一閱讀能力」，「卷二寫作能力」兩部分。文憑試沒有相關題目考察語音、詞彙及語法知識 (香港考試及評核局 2020)。與大陸高考語文全國卷相比，本書面語試卷在語音、詞彙、語法知識的檢測上，題型、題量和分值都較多。比如 2021 年高考全國統一考試「語文」，全卷 150 分，第三題是「語言文字應用」，分值 20 分，其中「語言文字運用 I」有三小題可算作檢測「語法能力」的題目，分別是選詞填空、選句填空、辨別病句，各佔 3 分，共 9 分。(中國教育在

線 2021)。香港的文憑試和高考都是常模參照測試 (norm-referenced test)，目的是對應試者的語文運用能力做出有效的區分，因而沒有或很少檢測語言知識的題目。而本試卷作為全面檢測書面中文能力的水平測試，對「語言能力」的評核則包括了語言知識和語言運用兩個方面，因此本試卷在着重檢測閱讀與寫作這兩種語言運用能力的同時，也就設置了相當數量的題目檢測語言知識。

效度 (validity) 指標，通俗地說，就是考試題目是否檢測了它所要檢測的內容。從表 1 可見，就內容效度 (content validity，即題目在多大程度上代表了它所要測量的目標) 而言，試卷的總體設計是合理的。

#### 四、試卷質量分析

##### 4.1 歷年試卷難度、區分度、信度數據的解釋與分析

本研究各項指標的計算方式為：難度 = 所有考生得分均值 / 該題滿分；區分度採取皮爾遜積差相關係數 (Pearson product-moment correlation coefficient)，即各題得分與總分的相關係數；信度採用克隆巴赫係數 (Cronbach's  $\alpha$ )。對於標準化考試，難度 0.3~0.7 之間屬於難度適中，小於 0.3 屬於偏難，大於 0.7 屬於偏易；區分度高於 0.3 就可以認為具有良好的區分度，信度高於 0.75 就可以認為試卷具有良好的信度 (Fulcher, 2010, pp. 182-183)。對於水平測試，由於考試成績會呈負偏分佈 (negatively skewed distribution) (Fulcher, 2010, p. 79)，也就是取得較好成績的考生佔多數，因此對這幾項指標的意義需作不同的解釋，難度係數大於 0.7 就不能簡單解讀為試題偏易，而要結合該試題檢驗的是何種能力進行分析。比如「選擇同音字」的歷年平均難度為 0.74，也就是說大部分考生在該題取得中上的成績，可以理解為該題難度是合適的，如果「選擇同音字」難度值過大，比如 0.9 或之上，很明顯這是不符合實際情況的，這時我們就應該認為該題屬於偏易的題目，需要加以改善。再如「區分度」體現的是該題目

表 2：歷年試卷的難度、區分度及信度

2010-2020 年度試卷難度、區分度、信度 <sup>1</sup>													
年份			2010	2011	2013	2014	2015	2016	2017	2018	2019	2020	均值
參與考試的人數			250	176	133	155	83	91	72	88	116	43	121
難度 (整體)			0.58	0.63	0.42	0.71	0.60	0.65	0.70	0.68	0.56	0.61	0.61
區分度 (整體)			0.28	0.28	0.30	0.19	0.24	0.21	0.24	0.23	0.25	0.21	0.24
試卷信度 Cronbach's Alpha (整體)			0.54	0.51	0.69	0.40	0.65	0.53	0.64	0.60	0.71	0.59	0.59
第一部分：用字辨析	1.1 選擇同音字 (2.5 分)	難度	0.64	0.96	0.68	0.70	0.45	0.80	0.91	0.96	0.61	0.67	0.74
		區分度	0.33	0.14	0.22	0.29	0.28	0.25	0.25	0.13	0.28	0.32	0.25
	1.2 給多音多義字注音 (2.5 分)	難度	0.38	0.32	0.59	0.57	0.60	0.43	0.64	0.41	0.58	0.42	0.49
		區分度	0.43	0.43	0.36	0.46	0.29	0.41	0.33	0.35	0.35	0.47	0.39
	1.3 辨別錯別字 (5 分)	難度	0.58	0.72	0.30	0.75	0.51	0.72	0.74	0.71	0.48	0.68	0.62
		區分度	0.23	0.23	0.24	0.14	0.19	0.17	0.22	0.30	0.25	0.16	0.21
第二部分：詞語辨析	2.1a 完成常用成語 (選擇題) (5 分)	難度	*	*	*	0.83	0.71	0.75	0.86	0.70	0.82	0.61	0.75
		區分度	*	*	*	0.11	0.17	0.28	0.23	0.30	0.11	0.07	0.18
	2.1b 完成常用成語 (自行填寫) (5 分)	難度	0.61	0.76	0.28	*	0.50	0.56	0.56	0.53	0.51	0.51	0.54
		區分度	0.27	0.30	0.45	*	0.32	0.27	0.31	0.24	0.40	0.17	0.30
	2.2 選詞填空 (5 分)	難度	0.62	0.72	0.36	0.73	0.81	0.74	0.79	0.75	0.67	0.73	0.69
		區分度	0.20	0.29	0.20	0.16	0.19	0.08	0.13	0.14	0.21	0.11	0.17
第三部分：辨別病句 (7 分)	難度	*	*	*	0.78	0.56	0.71	0.65	0.80	0.56	0.46	0.65	
	區分度	*	*	*	0.21	0.29	0.25	0.24	0.19	0.19	0.32	0.24	
第四部分：閱讀理解	4.1 閱讀文言文 (8 分)	難度	*	0.16	0.51	0.42	0.55	0.32	0.41	0.52	0.29	0.63	0.42
		區分度	*	0.29	0.26	0.16	0.27	0.16	0.29	0.25	0.31	0.14	0.24

\* 註：空白處表明由於該年的這個題幹曾做過修訂，因此該年的數據不適合加入表格做相關數據分析、特別是有關信度的分析。這包括：  
 第二部分：2.1a 完成常用成語 (選擇題模式) (2010、2011、2013 年)；  
 2.1b 完成常用成語 (自行填寫完成模式) (2014 年)  
 第三部分：辨別病句 (2010、2011、2013 年)  
 第四部分：4.1 閱讀文言文 (2010 年)

<sup>1</sup> 2012 年考試數據因故遺失。

與總分的相關性，如果區分度不高也應該結合難度及該題檢驗的語言能力加以分析，比如該題難度值低、區分度低，即無論書面中文能力高與低，該題的得分都偏低，這與水平測試成績負偏分佈的特點也是不一致的，那麼該題就可能偏難，需要改善。現將歷年試卷各題的難度、區分度及試卷整體信度的數據列表如下：

## 4.2 第一部分「用字辨析」分析

文字是書面語的載體，書面中文能力的測試，理應包括對漢字運用能力的檢測。漢字包括形音義三個方面。「用字辨析」共有三道大題，題型如下：

### 1.1 選擇同音字

說明：請選擇與劃線漢字的普通話讀音相同的漢字，並將答題紙相應的字母塗黑。

例：風雨同舟 A. 粥 B. 咒 C. 揍 D. 三個選項都不正確 (A)

### 1.2 給多音多義字注音

說明：請為下列劃線的漢字標注漢語拼音，將答案寫在答題紙相應的橫線上。

例：欽差大臣 (chāi)

### 1.3 辨別錯別字

說明：請審視下列語句劃線的漢字中有無錯別字。如有，請選出錯別字，把答題紙相應的字母塗黑，並把正確的字寫在後面的橫線上，如沒有，請將答題紙的 D 項塗黑。

例：女性的偉大幾乎不是顏費詩人所能明瞭的。(B / 廢)

A            B            C

1.1 檢測字音，1.2 檢測音與義，1.3 檢測字形。1.1 題要求考生從三個選項中選出劃線字在普通話裏的同音字，其難度均值 0.74，說明大部分考生都已可以正確辨別某字的普通話讀音。在普通話學習中，

知曉某字的普通話讀音與用普通話正確讀出該字相比，相對較易。此外，近年來，參加本校「普通話水平測試」的考生，平均分已經很接近二級乙等（80 分），這與 1.1 題的難度反映出考生辨音能力的情況是相近的，因為我們認為 1.1 的難度是合適的。後兩題的難度和區分度都適中。

## 4.3 第二部分「詞語辨析」分析

詞語是語言的建築材料，本部分的兩道題主要考察了考生對常用成語的掌握程度和辨析近義詞並選擇適合句子上下文的詞語的能力。題型如下：

### 2.1 完成常用成語

此題分兩個部分，一部分請學生選擇漢字完成成語，另一部分自行填寫合適的漢字，並把漢字寫在答題紙的橫線上。

### 2.2 選詞填空

此題給出幾段文字，略去了一些比較關鍵的詞語。請學生從方框中選擇最恰當的詞語，並把相應的字母填在劃線處，所列詞語不可重複使用。

2.1 完成常用成語的難度屬於合適的範圍，該題的區分度也相當良好；2.2 選詞填空的難度均值是 0.69，區分度也不高，說明能力高的考生與能力低的考生都可以在該題取得高分。該題用完形填空的形式，檢測在上下文語境中辨析詞語差別的能力，既涉及詞義的辨析，又涉及修辭，在有備選詞語供考生比較的情況下，大部分大學生應已具有這種能力。因而該題的質量是可以接受的。

## 4.4 第三部分「辨別病句」分析

句子是語言表達的基本單位，該檢測考生對存在語法、語義或用詞偏誤的句子的辨別能力。題型如下：

說明：請判斷下列句子的劃線部分是否符合現代漢語詞彙、語法規範。選出有錯誤的地方並將答題紙相應的字母塗黑，如果該句沒有錯誤，請將答題紙的 D 項塗黑。

例：不知你有無聽過，反正我長這麼大還是頭一回聽說這種事。  
(A)                      A                      B                      C

統計結果說明，該題難度適中，但區分度也不高，這說明大部分學生辨別病句的能力相近。

#### 4.5 第四部分「閱讀理解」分析

該部分包括文言文和現代文閱讀各一篇。閱讀文言文要求考生選擇常用文言字詞的意思，並回答有關問題。例如 2020 試卷閱讀的文章節選自《荀子·天論》，文中有「彊本而節用」一句，要求學生選擇的「本」的意思，選項有 A 基本、B 農業、C 商業；問答題比如「文中說：『君人者，隆禮尊賢而王，重法愛民而霸，好利多詐而危，權謀、傾覆、幽險而盡亡矣。』這句話對人君避免國家滅亡，提出了什麼建議？」表 2 顯示，文言文閱讀難度偏難，區分度也不高。這說明大部分學生文言文閱讀的能力不足，但是大家的不足也相近。

關於現代文閱讀，章熊認為，閱讀能力可以分為五類：認知能力（指對詞、句語義的辨析能力）、篩選能力（指從文字材料中捕捉關鍵性詞、語、句、段的能力）、闡釋能力（指把讀物中的詞語轉化為自己的語言的能力）、組合能力（指讀者通過綜合、歸納、調整，將從文本中獲取的資訊納入自己的知識和經驗系統，成為自己的認識的能力）、鑒賞、評價、創造能力（指評鑒文本的優劣、形成超越原讀物的獨特見解或思路的能力）（章熊，2000，頁 306-310）。現代文閱讀理解一題主要檢測的是後三種能力，因此題目均採取問答題的形式。要求考生對文章的主旨、文句蘊含的意思進行闡發或者文章的修辭技巧作出評價。例如 2014 年閱讀的文章是賈平凹的〈月跡〉，問題比如「縱觀全文，孩子們是如何追尋月的蹤跡的？」、「尋月的過程表現了

孩子們怎樣的天性？」，現代文閱讀歷年均在十年紀錄當中有八年都在七分以上，閱讀理解能力不俗。究其原因，是因為對於大學生來說，閱讀文獻（不僅是文學作品），進行分析並完成口頭或書面報告，是大部分課程學習的一種模式，因而，絕大部分的大學生具有良好的現代文閱讀理解、分析概括的能力，換句話說，該題目的各項係數真實地反映了考生的實際能力。

表 3：歷年閱讀理解（現代文）描述統計

年份	均值	中位數	標準差	偏度	最小值	最大值
2010	8.20	9.00	1.78	-1.54	0.00	10.00
2011	6.65	7.00	2.40	-0.38	0.50	10.00
2013	8.87	9.00	1.23	-4.84	0.00	9.50
2014	5.31	5.50	1.90	-0.37	0.00	9.50
2015	7.44	7.50	1.55	-0.75	2.00	10.00
2016	7.68	7.50	1.19	-0.37	5.00	10.00
2017	8.60	8.70	1.03	-0.66	6.00	10.00
2018	8.73	9.00	0.96	-0.81	6.00	10.00
2019	7.95	8.00	1.25	-1.51	1.50	10.00
2020	8.69	9.00	1.13	-0.63	6.00	10.00

#### 4.6 第五部分「寫作」分析

培養學生的寫作能力，是語文教育的基本目標之一。語文學習最實用性的兩個目標就是說得好、寫得好。題型要求學生從兩題中任選一題寫一篇文章。寫作可以說是最能展示考生書面中文能力的題目之一，因而在考卷分值上，本考試的寫作一題佔到了全卷總分的一半。為了深入了解這種情況，現將對歷年試卷寫作題的統計分析列表如下：

表 4：歷年寫作描述統計

年份	均值	中位數	標準差	偏度	最小值	最大值
2010	40.9	41.5	3.3	-1.1	24.7	47.4
2011	40.2	40.2	3.2	-0.42	30.2	48.5
2013	39.7	39.8	4.0	-0.18	27.5	47.8
2014	39.9	39.9	3.4	0.1	32.4	48
2015	38.9	38.6	3.7	0.02	31.5	46
2016	38.4	37.5	3.7	0.2	29.6	46.8
2017	38.4	38.5	3.3	0.01	29.6	46
2018	39.8	40.6	3.4	-0.8	29.6	47.8
2019	38.4	37.7	3.0	0.4	31.1	47.8
2020	42.2	41.9	2.3	0.4	38.8	47.5

為了比較客觀地評核作文的質量，我們在評分時將作文分為五級，並做出了清晰的等級描述。從上表所見，歷年「寫作」的平均分在 38 分到 42 分之間。對應作文的等級，屬於第三級和第四級，即考生可以寫出中心較為明確，結構較為完整、內容較為充實，語言較為通順的作文。所以，就此項而言，難度與作文的評分標準有直接關係。應該說，大部分大學生具有中等偏上的寫作水平，這是符合對大學生中文寫作能力的一般預期的，這也說明作文評分所定的標準是合理的。

#### 4.7 試卷總體信度的討論（首四部份，不包括閱讀現代文及寫作部份）

從表 2 可見，歷年試卷信度  $\alpha$  均值為 0.59。對於常模參照測試而言，這一信度值是偏低的。但是，由於水平測試成績缺乏變異性 (lack of variance)，這會導致 Cronbach's  $\alpha$  總是非常低 (always be very low) (Fulcher, 2010, p. 81)。教育測試家 Kane 指出，對於水平測試而言，只要信度  $\alpha$  不低於 0.5，試卷的質量就是良好的 (Kane, 1986)。這說明本試卷具有好的信度水平。

如想深入分析各題測試內容相關性，可利用相關性矩陣 (Inter-

Item Correlation Matrix) 檢視題目之間的關連性。以下我們再以 2020 年試卷為例，對各題測試內容的內部相關性分析如下：

表 5：2020 年試卷各題測試內容相關性

2020 年試卷內容項目相關性矩陣 Inter-Item Correlation Matrix <sup>3</sup>									
	選擇同音字	給多音多義字注音	辨別錯別字	完成常用成語	選詞填空	辨別病句	閱讀文言文	閱讀現代文	寫作
選擇同音字	1.00	0.50	0.39	0.10	0.20	0.29	-0.01	0.23	0.33
給多音多義字注音	0.50	1.00	0.35	0.18	0.12	0.52	0.20	0.22	0.16
辨別錯別字	0.39	0.35	1.00	0.18	0.29	0.42	0.28	0.37	0.22
完成常用成語	0.10	0.18	0.18	1.00	0.00	0.14	0.10	0.09	0.30
選詞填空	0.20	0.12	0.29	0.00	1.00	0.16	0.15	0.14	0.03
辨別病句	0.29	0.52	0.42	0.14	0.16	1.00	0.10	0.27	-0.06
閱讀文言文	-0.01	0.20	0.28	0.10	0.15	0.10	1.00	0.09	0.23
閱讀現代文	0.23	0.22	0.37	0.09	0.14	0.27	0.09	1.00	0.38
寫作	0.33	0.16	0.22	0.30	0.03	-0.06	0.23	0.38	1.00

如以「寫作」作為書面中文運用能力題的代表，上表數據顯示，「閱讀現代文」與其相關性最高 ( $r=0.38$ )，這也印證了「讀」與「寫」能力的密切關係，而「辨別病句」及「選詞填空」與「寫作」的相關性均很低。其他的題目，比如「選擇同音字」與「閱讀文言文」也顯示很低的相關性，「完成常用成語」與「選詞填空」之間是零相關。以上談及的幾個測試項目並非是嶺南的中文測試試卷獨有的，而是很多中文考試當中的常設項目，這說明試卷測試語言知識與語言能力的題目在性質上或存在着不少的差異。

對我們試卷未來的修訂來說，如何在「辨別病句」這類有關語言知識的題目中，體現語言運用的能力，是擬題時需要考慮的一個問題。此外，將試卷分為「卷一」與「卷二」，前者檢測書面語知識，後者檢測書面語運用能力，也是值得探討的一個設計方向。

<sup>3</sup> 該矩陣反映的是所有項目（題目）之間的相關性，結果在 0.15 至 0.5 之間，就可以認為相關性在合理的範圍之內。