

基礎華語文能力測驗試題分析

國家華語測驗推動工作委員會 曾文璇

摘要

本研究針對「國家華語測驗推動工作委員會」所研製的基礎級華語文能力測驗，根據預試試題之前後測統計值（難度、鑑別度），歸納出提高試題品質的修改方法。

一般華語測驗命題在設計上必須考量的原則有：一、題目是否符合自然真實的情況；二、題目所使用的主題情境是否為這個階段的學習者必須面對的；三、題目所給予的任務是否具有溝通功能；四、題目所給予的任務是否適合這一階段的學習者；五、整份試題所設計的主題情境與任務的給予是否均衡。

除了上述命題原則外，經研究人員大量審題與參考多次預試統計值，具體得出影響試題好壞的相關要點。基礎級測驗著重考生的「認知」與「溝通」能力，所以，審題時除了「語法、詞彙」使用適切性的考量外，「主題情境」也是審題的重點之一。

本研究將施測結果不理想的試題進行修改，重新進行施測，再經量化分析審視修改方法的適當性與成效，歸納修改的方法與原則。此外，希望透過前後測的結果找出影響試題品質不良的其他因素。

本研究希望透過基礎試題品質的提升，進而提高基礎級測驗的信度與效度，同時提供命題教師與測驗相關單位參考。

關鍵字：基礎華語文能力測驗、題目特徵、命題分析

一、前言

基礎測驗從 2005 年～2007 年共進行四份卷子海內外施測，預試人數接近七百多人次，研究人員依據預試完後的統計數值，從題目鑑別度、難度與「點二分析」¹，重新審視題目設計，研究人員將四份卷子 320 題，依其題目特徵分成四大類型，再將題目按題目特徵進行相關修改後，組成兩份卷子，於 2008 年進行二次施測。

兩份卷子裡，一份已施測完畢，另一份尚在進行中，預計 2009 年全部施測完畢。研究人員將施測完畢的卷子進行統計分析，數值顯示難度、鑑別度「點二分析」值皆提高。

本研究擬從 2007 年海外施測四份卷子與 2008 年已施測完畢的卷子，進行題目特徵分析，透過前後測的結果找出影響試題品質不良的其他因素。

二、題目特徵分析

基礎測驗所欲測驗能力為受試者「語言溝通能力」，測驗規劃題型與測試能力為：

表一 測驗題型

Part	聽力理解	閱讀理解
1	看圖回答	單句理解
2	問答理解	看圖釋義
3	對話理解	選詞填空
4	完成對話	完成段落

¹ 點二系列相關 (point-biserial correlation) 為相關分析的一種，適用於一個變項為二分名義變項 (考生答對或答錯某一題)，而另一變項為等距或比率變項 (考生的測驗總分)，分析二者之間的相關程度。在試題分析中可視為一種鑑別度的指標，數值愈高，表示試題愈具鑑別度，試題品質愈好。一般來說可以接受的數值在 0.2 以上。

表二 各大題測試能力表

Part	聽力理解	能力	閱讀理解	能力
1	看圖回答	字彙內在能力	單句理解	為資訊及論點而閱讀
2	問答理解	1.瞭解母語使用者間的交談 2.聽宣布及說明指導	看圖釋義	1.為資訊及論點而閱讀 2.導向閱讀
3	對話理解	瞭解母語使用者間的交談	選詞填空	文法內在能力(文法準確度)
4	完成對話	整體性聽力理解	完成段落	整體性的閱讀理解

基礎測驗題型設計涵蓋「分項測驗」與「整合式測驗」，規劃設計測出受試者在真實語境下的語文溝通能力²。今將 2006 年~2007 年在台灣預試的預試卷，進行量化統計分析，根據結果顯示，可將一份基礎卷 80 道題劃分四種類型：

- (一) 保留：無「零選項」現象，鑑別度大於 0.2，難度介於 0.6~0.85 之間，「點二系列相關」數值 >0.2，建議「保留」。
- (二) 小幅修改：無「零選項」現象，鑑別度介於 0~0.2 間，難度介於 0.85~0.95 之間，「點二系列相關」數值介於 0~0.2 之間，若鑑別度、「點二系列相關」有一個 <0.2，建議「修改」。
- (三) 大幅修改：具一個「零選項」，鑑別度 <0.2，難度 >0.95，「點二系列相關」數值 <0.2，建議「大幅修改」。
- (四) 修改或刪除：具兩個「零選項」，鑑別度為負值，難度 >0.95，「點二系列相關」數值為負值，建議「大幅修改或刪除」。

研究人員將四份卷子 320 道試題，依據上述四種題目特徵進行分類，分類結果顯示：

表三 分類結果

題型	數量	保留	小幅修改	大幅修改	修改或刪除
看圖回答		11	13	12	4
問答理解		16	13	9	2
對話理解		11	16	11	2
完成對話		24	10	4	2
單句理解		22	13	4	1
看圖釋義		22	4	13	1
選詞填空		22	10	6	2
完成段落		39	0	1	0

² 基礎測驗設計規劃，請參看 2008 年 ACTFL 會議論文 曾文璇〈從「語言溝通能力」探討 TOP-Huayu 基礎測驗〉。

由表三可知：

1. 「保留」題量：「完成段落」>「完成對話」>「選詞填空」、「看圖釋義」、「單句理解」。
2. 「小幅修改」題量：「對話理解」>「問答理解」、「單句理解」、「看圖回答」。
3. 「大幅修改」題量：「看圖釋義」>「看圖回答」>「對話理解」。
4. 「修改或刪除」題量：多集中在聽力理解部份，其中以「看圖回答」數量較多。
5. 「保留」數量：多集中在閱讀理解部份，其中以「完成段落」數量最多。
6. 「完成段落」鑑別度最佳；而「看圖回答」鑑別度最差。

關於第 4 點與第 5 點，研究人員假設在目標語區華語學習者因語言環境刺激影響，會加速其學習速度，對於聽力能力掌握度較高，基礎試卷所設計的聽力題對於這群受試者而言偏易。反之，非目標語區華語學習者，因學習環境中刺激較少，所以學習速度稍緩。將同樣的試題讓海外受試者受試，應該會產生「試題難度與鑑別度提高之結果」亦即，「保留」數量提高，而「修改或刪除」數量減少。若施測結果與此相符，則表示假設成立，亦可推論上述第 4、5 點，與題目設計無直接關係，而與受試者學習成果關係較大。

研究人員將上述四份卷中，取其中兩份至海外施測，施測國家包含美國紐約、洛杉磯、英國倫敦、法國巴黎、泰國、越南、日本以及韓國舉辦海外預試，共有 379 名考生參加。根據量化統計分析結果顯示，這兩份卷子的難度與鑑別度皆提高。

表四 量化統計分析結果

數量 題型	保留		小幅修改		大幅修改		修改或刪除	
	海外	國內	海外	國內	海外	國內	海外	國內
看圖回答	6	8	10	9	4	2	0	1
問答理解	11	11	7	4	1	4	1	1
對話理解	9	10	8	6	3	4	0	0
完成對話	18	14	2	3	0	0	0	3
單句理解	11	8	7	6	2	5	0	1
看圖釋義	9	9	4	7	6	4	1	0
選詞填空	11	10	4	7	5	1	0	2
完成段落	17	19	3	1	0	0	0	0
小計	92	89	45	43	21	20	2	8
	57.5%	55.6%	28.1%	26.8%	13.1%	12.5%	1.2%	5%

由表四得知：

- (1)「保留」題量，海外較國內多，兩者相差 1.9%。
- (2)「修改或刪除」題量，海外較國內少，兩者相差 3.8%。
- (3)「小幅修改」、「大幅修改」試題數量差異較少，「小幅修改」差異 1.3%；「大幅修改」差異 0.6%。

由結果得知，海外受測者因在非目標華語區學習華語，許多學習者使用華語的機會可能侷限於課堂內，而缺少環境刺激，所以，同樣的題目至海外施測，其施測結果證實難度、鑑別度與「點二系列相關」數值呈現較佳。雖然海外施測結果顯示難度、鑑別度與「點二系列相關」結果較佳，但「保留」題的數量增加 1.9%，而「修改或刪除」試題僅減少 3.8%。國內預試結果顯示共 71 題需要修改，而海外預試結果顯示共 68 題需要修改，在 160 道試題中，總修改題數僅減少 3 題，顯示整體試題設計偏易。

研究人員將 320 道試題重新審查，並進一步將同一類型題目集中，先從四個審題方向找出四種類型題目特徵。1.圖片設計合理性（清楚、真實、自然，無干擾）；2.題目設計合理性（符合邏輯、出題方向符合該大題設計）；3.選項、圖片、提問之間的關聯度（主題、情境範疇）；4.試題難度是否相符（組卷難度分佈比例）

表五 「保留」題特徵

共同特徵	<ol style="list-style-type: none"> 1.圖片設計清楚、簡單。 2.選項與提問之間關聯度高。 3.選項與提問屬於同一情境範疇。
分項特徵	<p>看圖回答</p> <ol style="list-style-type: none"> 1.選項詞彙關聯度極高，如：網球、棒球、籃球。 (A)他喜歡打棒球。 (B)他喜歡打網球。 (C)他喜歡打籃球。 <ol style="list-style-type: none"> 2.句式具變化性，並以簡單句型組織句式，如：「只」、「還有」等。 3.四個詞彙互相搭配於三個選項中，如 (A)她買了帽子和手套。 (B)她買了衣服和手套。 (C)她買了帽子和衣服。
	<p>問答理解</p> <ol style="list-style-type: none"> 1.對話中，使用句型，如「因為……所以」、「只」、「還」、「可是」等。 2.在回答句中含有拒絕、表達意願、說明原因、否定加以說明等漢語交際功能。
	<p>對話理解</p> <ol style="list-style-type: none"> 1.對話中，使用句型，如「因為……所以」、「只」、「還」、「可是」等。 2.在回答句中含有拒絕、表達意願、說明原因、否定加以說明等漢語交際功能。


		3.提問適切性。
完成對話		1.干擾選項與提問之間有一個以上詞彙相同。 2.正答選項與提問之間無完全相同詞彙。
單句理解		1.以句型帶入題幹。 2.在同一道題目中，不只一個資訊與論點詞彙。
看圖釋義		1.以句型帶入題幹。 2.在同一道題目中，不只一個資訊與論點詞彙。
選詞填空		1.同音字干擾選項，如「帶」與「戴」。 2.選項具同詞性，語法位置一致。 3.在教學上屬於同一組詞，如：才、再、就。
完成段落		因為本大題涉及篇章組織能力，屬於「整合能力」，所以需要受試者有足夠的語文知識，才能完成本大題。

表六 「小幅修改」試題特徵

共同特徵		1.圖片表達不清楚： 1.1 有一至二個干擾訊息。 1.2 訊息重點較不易辨識。 2.三個選項中有一個設計偏易。
分項特徵	看圖回答	1.選項詞彙屬同一主題情境：如打球、游泳、跳舞。 2.三個選項句式一致。 3.選項與提問皆為單一詞彙。 4.圖片中有選項或提問未涉及之詞彙。
	問答理解	1.對話中，使用句型，如「因為……所以」、「只」、「還」、「可是」等。 2.在回答句中含有拒絕、表達意願、說明原因、否定加以說明等漢語交際功能，但第二句話會直接陳述出訊息重點。
	對話理解	1.對話中，使用句型，如「因為……所以」、「只」、「還」、「可是」等。 2.在回答句中含有拒絕、表達意願、說明原因、否定加以說明等漢語交際功能，但言談最後一句話會直接陳述訊息重點。 3.提問適切。 4.有些題目圖片設計過於簡單，沒有呈現該對話情境。 在 2 與 4 的相互影響之下，造成鑑別度偏低。
	完成對話	1.干擾選項與提問之間有一個以上詞彙相同。 2.正答選項與提問之間無完全相同詞彙。
	單句理解	1.以句型帶入題幹。 2.在同一道題目中，只有一個資訊與論點詞彙。 3.圖片焦點模糊或過小，干擾受試者受測。
	看圖釋義	圖片焦點模糊或過小，干擾受試者受測。

	選詞填空	1. 選項詞彙主題差異過大，如：洗澡、洗碗、洗衣服。因為屬於閱讀題目，所以音近干擾較小。 2. 選項具同詞性。 3. 在語法教學上屬於同一組詞，如：才、再、就。
	完成段落	

表七 「大幅修改」試題特徵

共同特徵		1. 圖片表達不清楚： 1.1 有一至二個干擾訊息。 1.2 訊息重點不易清楚辨識。 1.3 高矮胖瘦等比較題目，無比較對象。 2. 三個選項中有兩個設計偏易。
分項特徵	看圖回答	三個選項屬同一簡單句式： 1. 選項詞彙屬分屬不同主題情境：如洗碗、洗澡、洗衣服。 2. 選項詞彙透過圖片可進行排除，亦即圖片具有暗示效果。如： Q: 王先生和誰一起坐計程車？ (A) 和他的太太。 (B) 和他的哥哥。 (C) 和他的妹妹。 
	問答理解	1. 對話中，幾乎不使用句型，九題中僅有一題使用句型「和」，其餘皆為直述句。 2. 在回答句中含有拒絕、表達意願、說明原因、否定加以說明等漢語交際功能，但第二句話會直接陳述出訊息重點。
	對話理解	1. 對話中，皆為直述句。 2. 在回答句中含有拒絕、表達意願、說明原因、否定加以說明等漢語交際功能，但言談最後一句話會直接陳述訊息重點。 3. 有些題目圖片設計過於簡單，沒有呈現該對話情境。 在 2 與 3 的相互影響之下，造成鑑別度偏低。
	完成對話	題幹設計為應酬語情境，如：你的鞋子很漂亮。 回應選項屬應酬語：謝謝、歡迎、沒關係、對不起、不客氣。
	單句理解	1. 四題中有兩題使用句型「都」；另外兩題圖片設計線條雜亂。 2. 在同一道題目中，只有一個資訊與論點詞彙。
	看圖釋義	1. 圖片焦點模糊或過小，干擾受試者受測。 2. 選項設計偏單詞理解與大題設計方向不符，如： (A) 屋子外面有鳥 (B) 屋子外面有貓 (C) 屋子外面有狗

	選詞填空	1.各選項詞彙主題差異過大。 2.選項設計偏單詞理解與大題設計方向不符。
	完成段落	僅有一題需要大幅修改，無法歸納試題特徵。

表八 「修改或刪除」試題特徵

共同特徵		1.圖片模糊不清。 2.除選項設計偏易，還有「可能選項」干擾。
分項特徵	看圖回答	1.選項互斥，如： Q：今天天氣怎麼樣？ (A)今天很熱 (B)今天很冷 (C)今天不冷也不熱 選項(A) (B)互斥。 2. 三個選項屬同一簡單句式。
	問答理解	兩題皆屬圖片干擾。
	對話理解	兩題皆屬圖片干擾。
	完成對話	選項設計為可能選項。
	單句理解	僅有一題，無法歸納試題特徵。
	看圖釋義	僅有一題，無法歸納試題特徵。
	選詞填空	選項設計偏單詞理解與大題設計方向不符。
	完成段落	

三、試題修改研究

研究人員接著依據「保留」試題特徵，進行試題修改，並於 2008 年進行海內外預試，預試總人數目前約兩百人左右。研究人員將預試結果進行量化統計分析，為求數值精確性，研究人員於「點二系列相關分析」中增加一項「與分測驗答對題數相關」分析，作為判斷試題鑑別度的參考指標之一。另依據研究人員內部進行一系列基礎試題與 CEFR 相關對應分析，分析結果顯示通過基礎測驗者，其聽力與閱讀的語文能力約相當於 CEFR A2 水準。研究人員進一步試圖從數值分析結果劃分試題等級，初步判定難度(P 值) 0.95，鑑別度(D 值) >0 ，且「點二系列相關」 >0.2 者，該題標記為 A1 水準；若難度 <0.6 ，則標記為「觀察題」，因為研究人員認為對受試者而言，此題難度水準可能已超過 A2 等級³。

³. 基礎試卷組卷由固定比例的 A1 題與 A2 題組成，題目排列由易至難。「點二系列相關」分析數值越高，顯示該題設計與該大題題型設計以及與整份測驗所預測得的能力具一致性。而試題等

此外，由於「完成段落」題目集中於「保留」題中，因此不做修改直接進行施測，施測結果亦顯示，此大題題目集中於「保留」。以下便針對其餘基礎測驗七大題型試題修改前後結果進行說明：

(一)看圖回答

1.試題特徵：

- (1)選項詞彙關聯度極高，如：網球、棒球、籃球。
- (2)句式具變化性，以簡單句型組織，如：「只」、「還有」等。
- (3)四個詞彙互相搭配於三個選項中。

2.試題修改方式：

修改方式(1)：將考點範疇縮小為細項詞彙，但不更動選項句式一致性。

2108006	通過率	難度	鑑別度	與答對總題數之 相關	與分測驗答對題數相關	建議
修改題	99.05%	0.983	0.034	0.213	0.276	A1
原題	100%	1	0		—	大幅修改

修改方式(2)：同樣的考點詞彙，但進行句型變化。

2108010	通過率	難度	鑑別度	與答對總題數之 相關	與分測驗答對題數相關	建議
修改題	42.86%	0.458	0.501	0.235	0.432	觀察
原題	100%	1	0		—	大幅修改

由修改方式(1)、(2)顯示，選項句式具一致性，對於此階段的受試者而言，題目的 P 值偏高；反之，選項句式富變化性，則題目 P 值又偏低。

(二)問答理解、對話理解

1.試題特徵

- (1)對話中，使用句型，如「因為……所以」、「只」、「還」、「可是」等。
- (2)在回答句中含有拒絕、表達意願、說明原因、否定加以說明等漢語交際功能。

2.試題修改方式

修改方式(1)：使用句型

級劃分增加「A1」題與「觀察」題，是考量整份卷子的難度應是「適中」。P 值、D 值「點二系列相關」數值不佳的題目，不宜出現，反之，題目偏難的「觀察題」，在組卷時亦需思考其出現比例。

2108020	通過率	難度	鑑別度	與答對總題數之 相關	與分測驗答對題數相關	建議
修改題	90.48%	0.858	0.2	0.223	0.403	保留
原題	95.19%	0.92	0.16		0.261	小幅修改

修改方式(2)：在回答句中含有拒絕、表達意願、說明原因、否定加以說明等漢語交際功能。

2108013	通過率	難度	鑑別度	與答對總題數之 相關	與分測驗答對題數相關	建議
修改題	92.38%	0.879	0.241	0.371	0.458	保留
原題	100%	1	0		0.001	大幅修改

「問答理解」與「對話理解」兩大題型，所預測得語言能力近乎相同，而「保留」試題特徵一樣，故以相同的方式進行修改，修改後的試題所呈現的數值顯示難度、鑑別度、「點二相關分析」較佳。

(三)完成對話

1.試題特徵

- (1) 干擾選項與提問之間有一個以上詞彙相同。
- (2) 正答選項與提問之間無完全相同詞彙。

2.試題修改方式

修改方式：干擾選項與提問之間有一個以上詞彙相同；正答選項與提問之間無完全相同詞彙。

2108035	通過率	難度	鑑別度	與答對總題數之 相關	與分測驗答對題數相關	建議
修改題	74.29%	0.793	0.414	0.264	0.443	保留
原題	94.27%	0.906	0.189		0.378	小幅修改

本大題設計無圖片、無選項，主要測驗受試者聽力整體能力。受試者需要較多的語言能力才能完成任務。題目經過修改後顯示通過率、難度、鑑別度與「點二分析值」較佳。修改方式，主要降低受試者遇到不會的題目進行「猜題」時的猜對比率。

(四)單句理解、看圖釋義

1.試題特徵

- (1) 以句型帶入題幹。

(2) 在同一道題目中，不只一個資訊與論點詞彙。

修改方式(1)：以句型帶入題幹。

2108043	通過率	難度	鑑別度	與答對總題數之 相關	與分測驗答對題數相關	建議
修改題	92.38%	0.897	0.207	0.457	0.458	保留
原題	97.73%	1	0		0.001	大幅修改

2108052	通過率	難度	鑑別度	與答對總題數之 相關	與分測驗答對題數相關	建議
修改題	82.86%	0.769	0.296	0.447	0.423	保留
原題	97.92%	0.962	0.075		0.186	大幅修改

修改方式(2)：在同一道題目中，不只一個資訊與論點詞彙。

2108045	通過率	難度	鑑別度	與答對總題數之 相關	與分測驗答對題數相關	建議
修改題	89.52%	0.879	0.241	0.45	0.506	保留
原題	77.54%	0.96	0.08		0.154	大幅修改

「單句理解」與「看圖釋義」兩大題型，所預測得語言能力近乎相同，而「保留」試題特徵一樣，故以相同的方式進行修改，修改後的試題所呈現的數值顯示難度、鑑別度、「點二相關分析」較佳。

(五) 選詞填空

1. 試題特徵

- (1) 同音字干擾選項，如「帶」與「戴」。
- (2) 選項具同詞性，語法位置一致。
- (3) 在教學上屬於同一組詞，如：才、再、就。

2. 修改方式：

修改方式(1)：選項具同詞性，語法位置一致。

2108063	通過率	難度	鑑別度	與答對總題數之 相關	與分測驗答對題數相關	建議
修改題	90.48%	0.858	0.2	0.418	0.445	保留
原題	97.73%	1	0		—	大幅修改

修改方式(2)：在教學上屬於同一組詞，如：才、再、就。

2108067	通過率	難度	鑑別度	與答對總題數之 相關	與分測驗答對題數相關	建議
修改題	70.48%	0.693	0.282	0.22	0.237	保留
原題	96.79%	0.96	0.08		0.239	小幅修改

本大題主要測驗受試者語法能力，在教學上屬於同一組詞，或是選項同詞性、語法位置相同，受試者除了知道詞彙在語法中如何被使用，還需進一步知道，語法的適用性，所以這兩種修改方式會提高試題難度。

四、結論

依據 2008 年施測結果顯示，整份試題題型設計對於受試者而言具一定程度的挑戰性，「完成段落」涉及篇章組織能力，屬於「整體性能力測驗」，對此階段受試者而言，題型整體設計偏難。「問答理解」、「對話理解」兩大題型設計涉及自然真實語境，其題目設計難點在於對話的邏輯性與真實性，而運用簡易句型，可提高試題難度。「單句理解」與「看圖釋義」兩大題型設計尋找資訊與論點能力。所以，增加資訊論點，可以提高試題難度。「完成對話」屬於「整體性能力測驗」，涉及語境整體性理解，並能做出適當回應。這一大題的選項設計是一大難點，但是，整體試題 P 值偏低，顯示對受試者而言此大題具一定難度。「看圖回答」屬於「單項能力測驗」中的「詞彙量」測驗，就受試者而言，屬於「語言知識內在能力」，這一大題題目較設計不易掌握，試題設計容易偏易或偏難。而「選詞填空」亦屬「語言知識內在能力」中的「文法內在能力」，這一大題設計偏重語法，所以，語法位置、使用方式等考法都可以有效提高難度。

華測會研發人員希望透過上述相關研究，提高基礎測驗試題的品質，進一步提升測驗本身的信效度。但是，此份研究仍待更多試題樣本，以檢核上述修改方式的適切性，並尋找更佳的試題修改方法，以使本研究更臻完備。

參考書目：

1. Council of Europe. 2001. *The Common European Framework of Reference for Languages: learning, teaching, assessment*.
2. 林清山《心理與教育統計學》(東華書局)，1992年。
3. 藍珮君〈基礎華語文能力測驗與歐洲語言共同參考架構的對應關係〉(2007年、第三屆華文教學國際論壇)。
4. 曾文璇〈從「語言溝通能力」探討 TOP-Huayu 基礎測驗〉(2008年，ACTFL 年會)。