

利用項目功能差異檢視澳門 電子閱讀素養表現的優勢和弱勢

薛寶嫦

電子閱讀評量 (Electronic Reading Assessment, 簡稱 ERA) 是「PISA 2009 閱讀素養測試計劃」中一種新型的電腦化評核方式。本文利用學生在電子閱讀先導測試的答案, 透過項目功能差異的分析方法檢視澳門十五歲學生於電子閱讀素養表現的優勢和弱勢。文中重點分析二十五個被鑑別出在澳門具有項目功能差異特徵的 ERA 測試項目, 以檢視學生在以下兩個方面的作答優勢和弱勢: (1)項目類型 (多項選擇題、提供答案題); (2)閱讀層面 (使用和提取、綜合和詮釋、反思和評估)。礙於試題和測試成績的保密原則, 筆者不能洩露項目內容和各國在先導測試各個項目所取得的成績, 只能利用 PISA 研究總部在 ERA 先導測試後所公佈的項目, 檢視其中內容, 以具體說明澳門十五歲學生電子閱讀素養表現成就及其不足之處。

關鍵字：閱讀素養、電子閱讀評量、項目功能差異、PISA

作者現職：澳門大學教育學院助理教授

壹、引言

學生能力國際評估計劃(Programme for International Student Assessment, 簡稱 PISA) 是由經濟合作及發展組織(Organization for Economic Co-operation and Development, 簡稱 OECD) 策劃每三年舉行一次的取樣調查, 主要分為閱讀素養、數學素養及科學素養三個領域, 旨在評估接近完成基礎教育的十五歲學生, 是否具備足夠的知識與技能以投身社會。PISA 2009 以閱讀素養為主要評估領域, 並新增電子閱讀評量(Electronic Reading Assessment, 簡稱 ERA) 為次要的評估領域, 以找出十五歲學生的網上閱讀水平(OECD, 1999 & 2007)。在 ERA 中, 閱讀任務一般會要求學生辨識重要的問題, 尋找與這項閱讀任務的預設存取路徑(access paths) 相配的訊息, 分析所提取訊息的用途, 從多份電子文本中綜合所提取訊息以回答問題, 以及透過電子的方法與他人交流和給予回覆(Cheung & Sit, 2008)。無可否認, 澳門學生對這種於新世紀數碼訊息時代出現的新型電子閱讀評量方式並不習慣(關於電子閱讀評量的發展趨勢和方法, 參閱葉雅鏞, 2008; Coiro & Dobler, 2007; Coiro, Knobel, Lankshear, & Leu, 2008; Katz, 2007; Rush, Eakle, & Berger, 2007; Sit & Cheung, 2009)。澳門語文教師也很少使用電子方法評核學生的閱讀理解水平, 因而並未在正式的課程文件中列為一般的評價慣例(Shan et al., 2007)。

貳、研究目的

如同其他學生學習表現的國際比較, 閱讀任務需要進行先導測試(pilot study) 以確定測試項目在不同的國家之間是以相同的方式表現。在今次的先導測試中, 首要考慮的是觀察那些擁有相同素養水平但來自不同國家的學生, 他們回答是否存在著顯著的差異。項目功能差異(differential item functioning, DIF) 是標準化測驗編製的一個重要環節, 也是目前國外許多大型測試題目(在本文中又稱為「項目」) 常規項目分析過程中一個不可或缺的部份(Holland & Wainer, 1993)。從心理測量的角度來說, 地方項目功能差異(country DIF, 意指某地方或國家與項目之相互作用) 的出現, 是指擁有相同素養水平, 但來自不同地方的答題者, 對某一個項目正確地回答有不同的機率(關於項目功能差異的評核方法, 可參閱 Penfield & Lam, 2000; 關於辨識有偏誤測試項目的方法, 可參閱 Camilli & Shepard, 1994; Clauser & Mazor, 1998; Yildirim & Berberoglu, 2009)。

本文目的是分析那些在 PISA 2009 先導測試中被辨識為有地方項目功能差異的 ERA 項目, 以揭示澳門十五歲學生於網上閱讀的優勢和弱勢。與其他 OECD 及參與計劃的國家/經濟體系(以下統稱為「地方」) 比較, 是有一些原

因會導致澳門學生回答一些項目時有項目功能差異，例如：學生對電子文本不一定熟悉、電子閱讀文本的翻譯不一定精準（關於高利害測驗測量建構的效度問題，參閱 Haladyna & Downing, 2004）等。對於參加今次先導測試且就讀於澳門中文學校的學生來說，他們可能會因為電子測試所預設使用的中文輸入法不熟悉，又或中文輸入速度太慢而影響閱讀素養表現。這些因素均有可能導致各地區學生的作答行為存在差異，特別是當學生需要回答提供答案型的項目時，預期這些因素的影響會較回答多項選擇的項目時明顯得多。

由於大多數先導測試的 ERA 項目沒有公開發佈，部分於 PISA 2009 正式測試（main survey）時更會再度使用。在為了保密而不透露項目內容的情況下，筆者較難具體地公開清楚說明導致某個項目有地方項目功能差異的原因，從而協助澳門語文教師如何在語文課程和教學上處理困惑和帶領學生走出誤區。儘管有這個限制，本文仍期望能利用 PISA 研究總部公開發佈的少數 ERA 項目（有兩個題組共七個項目）及相關的評分指引，闡明澳門十五歲學生在電子閱讀素養展現的優勢和弱勢。本文的研究數據來自就讀澳門中文學校十五歲學生於 2008 年四月所進行的先導測試，學生皆使用中文來回答測試項目，所有項目皆是根據國際英文版的項目翻譯而成。

叁、PISA 採用的地方項目功能差異的統計方法

本節扼要闡明 PISA 研究總部（PISA Consortium）如何引領各地方的 PISA 中心檢視 ERA 先導測試的作答數據，將各地方具有功能差異的項目辨別出來，然後綜合各地方那些沒有功能差異的項目，以便在正式測試中使用。雖然項目評分是由各地方的 PISA 中心負責完成，但項目功能差異分析卻是由 ERA 研究總部統一辦理。各地方的 PISA 中心只需利用 PISA 研究總部對各地方項目的項目功能差異分析結果，從中檢視出現項目功能差異的原因。文獻中，並不缺乏與項目功能差異有關的研究，然而 PISA 研究總部沒有採用較為專業的項目反應理論（item response theory）來辨別項目功能差異，而是採用較為容易理解的分析步驟，方便每一個參與測試的地方找出各自具有潛在功能差異的項目。從方法學的角度來說，這對於今次 ERA 先導測試各地方所收集的小量學生回應數據甚為適用。

ERA 有多項選擇（multiple choice）和提供答案（constructed response）兩種題型。其中，多項選擇題會再細分為簡單的多項選擇題和複雜的多項選擇題兩種，所謂複雜的多項選擇題是由多組的簡單多項選擇題組成，受試者需要答對全部的簡單多項選擇題，才能取得滿分的評分。提供答案題會再細分為封閉式提供答案題和開放式提供答案題兩種。不同的題型有不同的答對百分比

(facility) 計算方法。簡單的多項選擇題的答對百分比即答題通過率；複雜的多項選擇題、封閉式提供答案題和開放式提供答案題的答案則以部分給分 (partial credit scoring) 的方式評分，這些項目的答對百分比的計算方法如下：(1)若果項目最高得分 (maximum score) 是一分的話，答對百分比是指那些被評為獲得一分的學生的百分比；(2)若果項目最高得分是大於一分的話，答對百分比是各組分數答對百分比乘以這組對應分數的總和，再除以這項目的最高得分。

每一個 ERA 項目的項目功能差異指數，其計算步驟如下：(1)計算每一個項目的答對百分比，每一地方會分別計算；(2)計算每一地方每一個項目的「相對答對百分比」(national relative item facility)，方法是利用全部項目的答對百分比的算術平均值 (arithmetic mean)，計算每一個項目的答對百分比和這平均值的差；(3)計算每一個項目的「國際相對答對百分比」(international relative item facility)，方法是將第二步的每一地方各項目的「相對答對百分比」加總，然後取算術平均。最後是計算每一地方各項目「相對答對百分比」和「國際相對答對百分比」的差，即為項目功能差異指數。須留意的是，上述步驟需使用其他地方的數據計算項目功能差異指數，所以指數是會受到其他地方數據的影響而有差異，因此，為表明指數取值的可能範圍，指數會以區間表示。再者，項目功能差異指數可以根據正負值分為正向和負向兩種，以下利用兩條公佈的項目為例加以說明。

澳門數據顯示，E010Q02 (網絡詐騙) 是一個容易的項目，有 81% 的通過率 (見表 1)。E022Q08 (大家講) 是一個困難的項目，有 14% 的通過率。在比較「相對答對百分比」和「國際相對答對百分比」的差異後，E010Q02 的項目功能差異指數為正向 15-20%，而 E022Q08 則為負向 10-15%。根據 PISA 的內部指引，若指數小於 10%，該項目會被判定為不具備項目功能差異；若指數在 10% 至 15% 之間，該項目會被判定為有中等項目功能差異；若指數大於 15%，該項目會被判定為有高度項目功能差異。因此，E010Q02 在澳門被判定是一條有高度地方項目功能差異的項目，而 E022Q08 則被判定是一條有中等地方項目功能差異的項目。本文正是利用各項目所展示的不同程度的正向和負向的項目功能差異，按題目類型和閱讀層面對 ERA 題目進行細目分析，以闡明澳門十五歲學生在電子閱讀素養展現的優勢和弱勢。

上述是 ERA 項目在先導測試後得以優化的關鍵步驟，所有被判定有正向和負向地方項目功能差異的項目，未經修訂是不會用於正式測試中，從而保證參與測試的地方能對學生的電子閱讀素養有公平的國際比較 (Bond, Moss, & Carr, 1996)。

肆、先導測試中具有項目功能差異的 ERA 項目的特徵

在澳門 PISA 2009 先導測試中被判定為具有項目功能差異的 ERA 項目總共有二十五個（見表 1）。鑑於測試項目必須保密的緣故，只有那些在先導測試完結後由 PISA 研究總部公佈的項目（即「網絡詐騙」和「大家講」兩個題組）才會在本文的表中列出。澳門十五歲學生在二十五個 ERA 項目的答對百分比由 0% 至 81% 不等，其中 E010Q02 是學生最感容易的項目，而 E010Q07T 則是最感困難的，這是一條由七組簡單的多項選擇題組成的複雜多項選擇題，沒有一位澳門十五歲學生能七組全部答對，因而沒有學生能取得滿分的評分。根據奧地利、比利時、加拿大、哥倫比亞、德國、丹麥、法國、中國香港、匈牙利、冰島、愛爾蘭、日本、韓國、中國澳門、挪威、波蘭、西班牙和瑞典等地方學生作答的數據，可以利用前述方法，計算澳門的每一個項目的項目功能差異指數的數值。

在澳門，共有十二個項目被判定為擁有正向的地方項目功能差異，顯示跟各個參與 ERA 測試地方的電子閱讀素養表現處於同一水平的澳門學生，他們有較高的機率答對問題。另一方面，共有十三個項目被判定為擁有負向的地方項目功能差異，顯示跟各地方的電子閱讀素養表現處於同一水平的澳門學生，他們有較低的機率答對問題。在 PISA 2009 ERA 先導測試中，各項目屬於兩種主要類型：(1)多項選擇題（又分為簡單和複雜兩種）；(2)提供答案題（又分為開放式和封閉式兩種）。這些項目的閱讀層面則有三種主要類型：(1)使用和提取；(2)綜合和詮釋；(3)反思和評估。本文正是針對各項目的兩款出題特色（即題型、閱讀層面）檢視澳門十五歲學生閱讀素養表現的優勢和弱勢。

表 1 澳門 PISA 2009 先導測試中具有項目功能差異的 ERA 項目的特徵

項目編號	答對百分比 (%)	項目功能差異指數 *	題型	閱讀層面
E002Q02	53	~35% (+)	封閉式 提供答案題	使用和提取
E002Q04	42	15-20% (-)	開放式 提供答案題	綜合和詮釋
E005Q04	13	20-25% (-)	開放式 提供答案題	反思和評估
E006Q05	4	10-15% (-)	封閉式 提供答案題	反思和評估

續表 1 澳門 PISA 2009 先導測試中具有項目功能差異的 ERA 項目的特徵

項目編號	答對 百分比 (%)	項目功能差 異指數* (+/-)	題型	閱讀層面
E007Q02	79	10-15% (+)	多項選擇題	使用和提取
E007Q03	80	~20% (+)	多項選擇題	綜合和詮釋
E007Q06	41	25-30% (+)	多項選擇題	綜合和詮釋
E008Q02T	79	15-20% (+)	多項選擇題	綜合和詮釋
E008Q03T	41	15-20% (+)	多項選擇題	綜合和詮釋
E008Q04T	54	25-30% (+)	多項選擇題	反思和評估
E010Q02 (網絡詐騙)**	81	15-20% (+)	多項選擇題	使用和提取
E010Q04 (網絡詐騙)**	75	15-20% (+)	多項選擇題	使用和提取
E010Q07T	0	~25% (-)	複雜的 多項選擇題	綜合和詮釋
E013Q01	45	10-15% (-)	多項選擇題	綜合和詮釋
E013Q03	72	10-15% (+)	多項選擇題	綜合和詮釋
E014Q05	74	20-25% (+)	多項選擇題	使用和提取
E014Q07	64	10-15% (+)	多項選擇題	綜合和詮釋
E014Q10	26	10-15% (-)	多項選擇題	綜合和詮釋
E017Q01	37	10-15% (-)	多項選擇題	綜合和詮釋
E017Q03	7	~20% (-)	多項選擇題	綜合和詮釋
E017Q06	24	25-30% (-)	開放式 提供答案題	反思和評估
E021Q06	24	10-15% (-)	多項選擇題	綜合和詮釋
E022Q02	34	20-25% (-)	多項選擇題	綜合和詮釋
E022Q08 (大家講)**	14	10-15% (-)	多項選擇題	綜合和詮釋
E022Q09 (大家講)**	15	10-15% (-)	開放式 提供答案題	反思和評估

* 數據是基於以下參與 PISA 2009 先導測試的國家 / 經濟體系整理而成：奧地利、比利時、加拿大、哥倫比亞、德國、丹麥、法國、中國香港、匈牙利、冰島、愛爾蘭、日本、韓國、中國澳門、挪威、波蘭、西班牙和瑞典。

** 基於保密原則，只有那些在先導測試之後由 PISA 研究總部公佈的項目，其項目名稱才會在本表中列出。

表 2 載有上述二十五個在澳門 PISA 2009 先導測試中被判定為具有項目功能差異的 ERA 項目，這些項目按四個維度進一步交互劃分：(1)題目類型；(2)閱讀層面；(3)題目難度；和(4)項目功能差異方向。在這二十五個項目當中，只有四個再度入選為 PISA 2009 ERA 正式測試項目，其中 E014Q07 (中等難度水平、正向項目功能差異)是澳門學生相對有優勢的項目；而 E013Q01、E017Q01

利用項目功能差異檢視澳門電子閱讀素養表現的優勢和弱勢

和 E006Q05（中等至高難度水平、負向項目功能差異）是澳門學生相對處於弱勢的項目。顯然易見，項目功能差異方向，無論正負，均有需要查找原因，修正項目後才能成為正式測試項目，以期符合效度評鑑的要求。

表 2 根據項目特徵和難度對具有項目功能差異的 ERA 項目的劃分結果

項目特徵	具有項目功能差異的 ERA 項目*			
	容易	中等難度		困難
	項目功能差異 (正向 +)	項目功能差異 (正向 +)	項目功能差異 (負向 -)	項目功能差異 (負向 -)
多項選擇題				
1. 使用和提取	E007Q02 E010Q02 (網絡詐騙) E010Q04 (網絡詐騙) E014Q05			
2. 綜合和詮釋	E007Q03 E008Q02T E013Q03	E007Q06 E008Q03T E014Q07 [#]	E013Q01 [#] E017Q01 [#] E022Q02	E010Q07T E014Q10 E017Q03 E021Q06 E022Q08 (大家講)
3. 反思和評估		E008Q04T		
提供答案題				
1. 使用和提取		E002Q02		
2. 綜合和詮釋			E002Q04	
3. 反思和評估				E005Q04 E006Q05 [#] E017Q06 E022Q09 (大家講)

* 澳門十五歲學生答對百分比大於 70 % 的項目被判定為容易，小於 30 % 被判定為困難，而 30 - 70 % 之間則屬中等難度水平。

入選 PISA 2009 正式測試但被鑑別出對澳門有項目功能差異的 ERA 項目。

表 2 的第一個研究結果是：在十二個被判定為具有正向的地方項目功能差異項目當中，十個顯示澳門的十五歲學生與參加是項國際測試同等素養水平的學生比較，主要優勢是在「使用和提取」和「綜合和詮釋」兩個閱讀層面，特

別是以多項選擇這種題目類型考核學生的項目。

表 2 的第二個研究結果是：在十三個被判定為具有負向的地方項目功能差異項目當中，全部顯示澳門的十五歲學生與參加是項國際測試同等素養水平的學生比較，主要弱勢是在「綜合和詮釋」和「反思和評估」兩個閱讀層面。在「綜合和詮釋」的閱讀層面中，弱勢的項目類型多數是多項選擇；在「反思和評估」的閱讀層面中，弱勢的項目類型全是提供答案型。

伍、先導測試中具有正向項目功能差異的 ERA 項目所揭示的澳門學生優勢

表 3 總結出澳門 PISA 2009 先導測試中具有正向項目功能差異的 ERA 項目，這些項目所揭示的澳門十五歲學生優勢以多項選擇題的作答方式為主，絕大部分集中在「使用和提取」和「綜合和詮釋」兩個閱讀層面。

在「使用和提取」閱讀層面，學生的優勢具體表現在懂得從眾多資料來源中使用和提取特定的資訊，例如：(1)從網頁文本中提取明確且直接闡述的定義；(2)將相同意思但不同文本排版方式的資料配對；(3)將某 URL 網址和一些給定的網頁連結；以及(4)識別貼切有用的搜尋結果，而這些結果是需要利用不太明顯但明確界定的關鍵詞。

在「綜合和詮釋」閱讀層面，學生的優勢具體表現在對文本形成廣泛的理解和針對閱讀目的作出恰當的詮釋，例如：(1)從一篇複雜且不熟悉的文本中總結要點；(2)從簡單的描述性的字詞中確認主題，以便識別貼切有用的搜尋結果；(3)將兩篇文本的資訊綜合，以便詮釋觀點是否雷同；(4)比較兩個網站所載資料，以便區別主題內容是否有異；(5)辨識一幅統計圖表的發展趨勢。

最後，在「反思和評估」閱讀層面的優勢不甚明顯，只有少許跡象顯示學生懂得評鑑從網站搜尋得到結果，判定資料是否貼切適用，是否準確無誤而值得信賴。

表 3 澳門 PISA 2009 先導測試中具有項目功能差異的 ERA 項目所揭示的優勢

問題編號	題型	題旨
使用和提取		
E007Q02	多項選擇題	提取訊息
E010Q02 (網絡詐騙)	多項選擇題	提取訊息
E010Q04 (網絡詐騙)	多項選擇題	提取訊息
E014Q05	多項選擇題	提取訊息
E002Q02	提供答案題	提取訊息
綜合和詮釋		
E007Q03	多項選擇題	形成廣泛理解
E007Q06	多項選擇題	發展詮釋
E008Q02T	多項選擇題	發展詮釋
E008Q03T	多項選擇題	發展詮釋
E013Q03	多項選擇題	發展詮釋
E014Q07 [#]	多項選擇題	發展詮釋
反思和評估		
E008Q04T	多項選擇題	反思和評估文本的內容

[#] 入選 PISA 2009 正式測試但被鑑別出對澳門具有項目功能差異的 ERA 項目。

PISA 2009 ERA 先導測試後透露 PISA 研究總部公佈的項目內容是容許的。從「網絡詐騙」這一題組的兩個公佈項目(即 E010Q02 和 E010Q04)顯示,無論資訊是在同一或不同網頁,澳門十五歲學生精於點擊超連結查找所需要的網上資料。網絡詐騙似乎是學生非常熟悉的課題,學生有能力勝任和樂於完成測試任務(參閱附錄 1 和 2 的項目內容、題旨、作答路徑和評分指引)。

值得注意的是:E014Q07 是一個被辨識為有正向項目功能差異的項目,因此這個項目有需要重新修訂,然後才能入選在 PISA 2009 ERA 正式測試項目當中。經澳門 PISA 測試中心審慎評估,認為這個項目輕微修改後是會適用於澳門。在不能夠透露這個項目的具體內容情況下,只能透露這是一個要求學生從圖表中辨識趨勢的項目,在日常的數學課中,澳門學生經常會遇上這一種學習任務的。

陸、先導測試中具有負向項目功能差異的 ERA 項目所揭示澳門學生的弱勢

表 4 總結出澳門 PISA 2009 先導測試中具有負向項目功能差異的 ERA 項目,這些項目所揭示的澳門十五歲學生弱勢全部集中在「綜合和詮釋」和「反思和評估」這兩個閱讀層面。「綜合和詮釋」項目的作答方式,絕大多數是多項

選擇題，而「反思和評估」項目的作答方式，全部屬於提供答案題型。

在「綜合和詮釋」閱讀層面，學生的弱勢具體表現在未能認識各種電子文本的功能和作用，例如：(1)將一連串往來的電郵訊息綜合，形成一個與電郵所陳述的要求相互吻合的建議；(2)認出哪些建議是文本中所提及和贊成的；(3)應用某一文本所述及的標準，並據此闡釋另一文本的內容和形式；(4)將不同網站所載內容綜合出共同傳達的訊息；(5)從一系列簡短文本中辨識出共同的課題；(6)在複雜的技術性陳述中對某一問題作出精要的解釋；(7)從一系列廣告中知悉某項服務的有用資訊；(8)從某個簡短電郵中洞悉其撰寫動機。

在「反思和評估」閱讀層面，學生的弱勢具體表現在未能從電子文本的內容或形式評鑑網頁的資料是否確實無誤，以及判定網站是否可資信賴。學生作出反思和評估時，有需要引用以前習得的知識，結合從電子文本提取的資料，提出有根有據的推理分析，來支持自己所作出的主張。

表 4 澳門 PISA 2009 先導測試中具有項目功能差異的 ERA 項目所揭示的弱勢

問題編號	題型	題旨
綜合和詮釋		
E010Q07T	多項選擇題	發展詮釋
E013Q01 [#]	多項選擇題	發展詮釋
E014Q10	多項選擇題	發展詮釋
E017Q01 [#]	多項選擇題	形成廣泛理解
E017Q03	多項選擇題	發展詮釋
E021Q06	多項選擇題	發展詮釋
E022Q02	多項選擇題	發展詮釋
E022Q08 (大家講)	多項選擇題	發展詮釋
E002Q04	提供答案題	發展詮釋
反思和評估		
E005Q04	提供答案題	反思和評估文本的內容
E006Q05 [#]	提供答案題	反思和評估文本的內容或形式
E017Q06	提供答案題	反思和評估文本的內容
E022Q09 (大家講)	提供答案題	反思和評估文本的內容

[#] 入選 PISA 2009 正式測試但被鑑別出對澳門具有項目功能差異的 ERA 項目。

從「大家講」這一題組的兩個公佈項目（即 E022Q08 和 E022Q09）顯示，澳門十五歲學生並不精於發展詮釋，亦不慣於對閱讀的文本進行反思和評估。E022Q08 要求學生識別文本提出了哪一套建議，而 E022Q09 則要求學生選取權威文章，結合已有知識和文章中的訊息，解釋為何該文較其他文章可靠。澳門學生在這兩個項目的表現，均較其他地方相同素養水平的學生為弱（參閱附錄

3 和 4 的項目內容及題旨、作答路徑和評分指引)。

顯然易見的是，澳門的初中學生較為習慣回答以滑鼠點擊選取答案的多項選擇題，多於以中文輸入法鍵入回答的提供答案題。當要回答的提供答案題需要對閱讀的材料反思和評估時，學生一般未能取得滿分，最多只能獲得部分分數。這現象很有可能是由於在澳門日常的測驗和考試中，強調的是對閱讀材料的正確理解，學生卻很少對閱讀內容提出質疑和爭辯。

最後，有三個被辨識為有負向項目功能差異的項目（即：E013Q01、E017Q01、E006Q05）仍然入選 PISA 2009 ERA 正式測試之中。這些項目經澳門 PISA 測試中心審閱，認為輕微修改後是適用於澳門。在不能夠透露這個項目的具體內容情況下，只能透露 E017Q01、E006Q05 這兩個項目的措詞未夠精準，有礙學生對項目達致正確和全面的理解，以進行綜合思考、反思和評估。至於 E013Q01 則要求學生綜合一系列的訊息，形成一個與訊息要求一致的意見。內容分析顯示澳門學生可能並不精於這一類以問題為本的邏輯思考。有趣的是，此現象亦出現於另一個使用中文的經濟體系。由於此項目會被保留於 PISA 2009 正式測試中，澳門 PISA 測試中心會密切注視正式測試結果會否與先導測試相同。

柒、研究限制

本研究共有兩個限制。首先，測試項目除非在先導測試後由 PISA 研究總部公開發佈，否則測試項目的內容是禁止透露的。因此，研究結果的討論，只能局限在與評核框架有關的內容，例如題目類型和閱讀層面。其次，除非得到其他參與國家的授權，否則其他國家的評核結果亦是禁止透露的。基於這兩個限制，本研究所揭示的正向和負向的項目功能差異不能與其他與澳門有類似教育背景的地方（例如香港）進行比較討論。大量的 ERA 測試項目被辨別為在澳門有項目功能差異，然而筆者不可能透露實質內容，因為這些項目或者是有待公開發佈，又或者保留供 PISA2009 正式測試或將來使用。

捌、結論

運用項目功能差異是 ERA 題目在先導測試後得以優化的關鍵步驟，保證參與測試的地方能對學生素養有公平的國際比較。在這個優化的過程中，各地方的 PISA 中心得知學生於先導測試的表現，初步了解學生的優勢和弱勢所在。

為達致公平和有效的國際比較，PISA2009 電子閱讀素養先導測試在澳門被辨識為具有項目功能差異的二十五個項目中的大多數已排除在 PISA 2009 正式測試之外，餘下的少部份澳門 PISA 測試中心會再作修訂，然後才予保留在正式測試中，因此 PISA 2009 正式測試的項目應不受項目功能差異影響。由是之故，是次揭示澳門十五歲學生的優勢與弱勢的研究設計並不適用於正式測試所收集得來的數據。

於先導測試運用項目功能差異分析方法，研究發現澳門十五歲學生與參與是項國際測試同等素養水平的其他地方學生比較，主要優勢是在「使用和提取」的閱讀層面，特別是以多項選擇題目類型考核學生的項目；主要弱勢是在「反思和評估」的閱讀層面，全部屬於提供答案題型。而「綜合和詮釋」的閱讀層面，既是澳門學生的優勢，也是澳門學生的弱勢。這些具優勢和弱勢具項目功能差異的題目同是多項選擇題，因此有需要繼續進行探討，才能得出確切的結論，用以幫助澳門學生發揮優勢，克服弱勢。

參考文獻

- 葉雅鏞 (2008)。國中學生線上閱讀素養與策略調整特徵之探討。國立臺南大學測驗統計研究所碩士論文，未出版，臺南。
- Bond, L., Moss, P., & Carr, P. (1996). Fairness in large-scale performance assessment. In G. W. Phillips (Ed.), *Technical issues in large-scale performance assessment* (pp. 117-140). Retrieved 3 December, 2009, from <http://nces.ed.gov/pubs/96802.pdf>
- Camilli, G., & Shepard, L. (1994). *Methods for identifying biased test items (Vol. 4)*. London: Sage.
- Cheung, K. C., & Sit, P. S. (2008). Electronic reading assessment: The PISA approach for the international comparison of reading comprehension. *Journal of Educational Research and Development*, 4(4), 19-40.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31-44.

- Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, 42(2), 214-257.
- Coiro, J., Knobel, M., Lankshear, C., & Leu, D. (2008). *The handbook of research on new literacies*. UK: Routledge.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Katz, I. R. (2007). Testing information literacy in digital environments: ETS's iSkills assessment. *Information Technology and Libraries*, 26(3), 3-12.
- OECD (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: Author.
- OECD (2007). *PISA 2009 reading framework – 23rd meeting of the PISA Governing Board, Oslo, 12-14 March, 2007*. Paris: Author.
- Penfield, R. D., & Lam, T. C. M. (2000). Assessing differential item functioning in performance assessment. *Educational Measurement: Issues and Practice*, 19(3), 5-15.
- Rush, L. S., Eakle, A. J., & Berger, A. (Eds.). (2007). *Secondary school literacy: What research reveals for classroom practice*. Illinois, US: National Council of Teachers of English.
- Shan, W. J., Jeong, S. L., Cheung, K. C., Zhang, C. L., Tian, Y., Wang, P. C., Vong, S.K., Tse, K. S., Leung, S. O., & Sze, T. M. (2007). *An investigatory study of the non-high education curriculum and its direction for reform in Macao: Final report*. Macao: DSEJ.
- Sit, P. S., & Cheung, K. C. (2009, September). *An examination of the design and implementation logistics of electronic reading assessment in PISA 2009 study: The Macao experience*. Paper presented at the PISA Research Conference, Kiel, Germany.

專論

Yildirim, H. H., & Berberoglu, G. (2009). Judgmental and statistical DIF analyses of the PISA-2003 mathematics literacy items. *International Journal of Testing*, 9(2), 108-121.

附錄1 E010Q02 (網絡詐騙)



網絡詐騙：題目 1 E010Q02

你在網絡詐騙網上資源站的首頁。根據這個網頁，以下哪一項是詐騙電郵的特徵？

A. 它要求個人資料。
B. 它包含不須要的廣告。
C. 它提供真正的服務。
D. 它來自有名的公司。

題旨：
使用和提取：提取資訊。
尋找一個明確說明的定義的重要的組成部分。

作答路徑：
根據「網絡詐騙資料網站」所載資料提取訊息以正確回答多項選擇題。
(註：虛線箭頭指向答案的所在位置，考生毋需導航至另一網頁以取得答案)

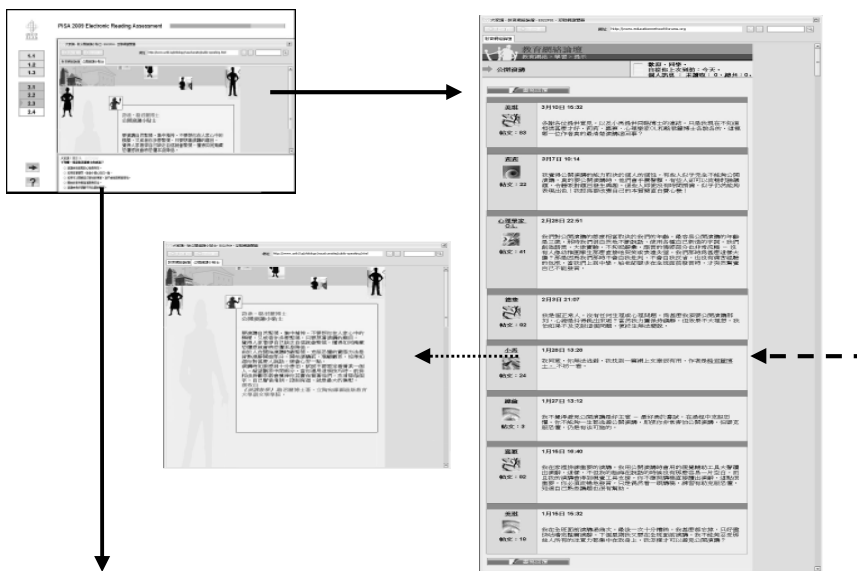
評分指引：
滿分
代碼 1： A. 它要求個人資料。
零分
代碼 0： 其他回答。
代碼 9： 沒有作答。

附錄 2 E010Q01 (網絡詐騙)



網絡詐騙：題目 2	E010Q01
全球寄出的網絡詐騙電郵平均一個月有多少封？	
A. 1,200。 B. 超過 60 億。 C. 大約 25,000。 D. 55,000。	
題旨： 使用和提取：提取資訊。 辨析一個目錄中提及的數字。	
作答路徑： 根據「網絡詐騙資料網站」所載資料提取訊息，以正確回答多項選擇題。(註：虛線箭頭指向答案的所在位置，考生毋需導航至另一網頁以取得答案)	
評分指引： 滿分 代碼 1： B. 超過 60 億。 零分 代碼 0： 其他回答。 代碼 9： 沒有作答。	

附錄 3 E022Q08 (大家講)



大家講：題目 3

E022Q08

下列哪一項是駱君麗博士的建議？

- A. 演講時抱輕鬆的心態最有效。
- B. 如果想著觀眾，就會少擔心自己一點。
- C. 如果可以隱藏自己害怕的事實，就不會覺得那麼害怕。
- D. 開始前背熟整篇演辭最安全。
- E. 演講時最好環顧不同位置的觀眾。

題旨：

綜合和詮釋：發展詮釋。

識別文本提出了哪一套建議。

作答路徑：

點擊「教育網絡論壇」網頁上的超連結（粗虛線箭頭指向超連結的所在位置） 前往所需網頁（幼虛線箭頭所指位置）。根據展示資料綜合一組駱君麗博士的提議，以正確回答多項選擇題。

評分指引：

滿分

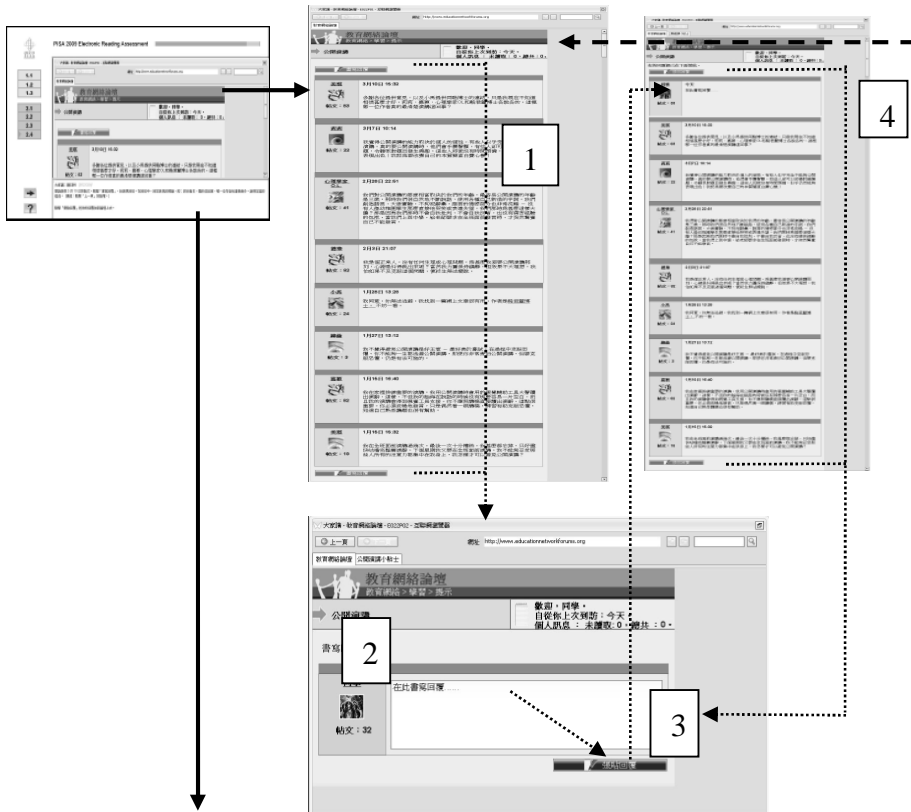
代碼 1： C. 如果可以隱藏自己害怕的事實，就不會覺得那麼害怕。

零分

代碼 0： 其他回答。

代碼 9： 沒有作答。

附錄 4 E022Q09 (大家講)



大家講：題目 4 E022Q09

閱讀美琪 3 月 10 日的貼文。點擊「書寫回覆」，給美琪回信。在回信中，回答美琪的問題，即：按你看來，關於這回事，哪一位作者知道得最多。說明答案的理由。【備註：點擊「上一頁」到論壇。】

點擊「張貼回覆」把你的回覆加到論壇上去。

題旨：

反思和評估：反思和評估文章的內容。

選取權威文章，結合已有知識和文章中的訊息，解釋為何該文較其他文章可靠。

作答路徑：

點擊位於「教育網絡論壇」網頁頂部或底部的「書寫回覆」按鍵（第一步），鍵入一個給美琪的回覆（第二步）。然後點擊「張貼回覆」把回覆加到論壇上（第三步）。如有需要，可以點擊位於網頁頂部或底部的「修改回覆」按鍵編輯答案（第四步）。（註：粗虛線箭頭指向開始答題的位置，而一組順序排列的幼虛線箭頭指出設計者對於連結網頁所預設的作答路徑）

評分指引：

滿分

代碼 1：確認駱君麗博士和 / 或心理學家 O.L.（明示或暗示），並且指出他們的專業資格。也可以表達對他們專業資格的懷疑。

- 兩位專家對此了解最多，但只有駱博士給予了解決問題的建議。
- 心理學家 O.L.或駱君麗博士，因為他們二人受訓於此範圍。
- 駱君麗博士。這篇貼文是唯一有大學背景的貼文。
- 駱君麗博士，只因她來自大學。
- 大學教授有豐富的公開演說經驗。
- 小馬已經對此調查並且發現一篇由懂得怎樣做的人寫的文章。這個男子在這種事情上顯而易見是一名專家，因此我認爲你應該聽從他的建議。
- 我最重視寫書的那一位，只因她出版了一本關於這論題的書。
- 心理學家 O.L.聽似權威，但當然你不會知道她是否真的是心理學家。
- 你應該跟隨心理學家 O.L.的建議，因為不僅他是一個有經驗的心理學家，而且他準確回答關於公開的講話的全部問題，並且非常可信。
- 心理學家 O.L.。因為他是一個受過訓練的心理學家。
- 或許對這最了解的人是博士。他是最有經驗的（或者至少比莉莉或者嘉穎有經驗），因此我認爲他更可信。

確認美琪提及的四位作者中的一位（莉莉、嘉穎、心理學家 O.L. 或駱君麗博士），並且提供一個和文章一致的理由，更兼提供一個關於貼文是有說服力、中肯或邏輯的理由。

- 心理學家 O.L.，只因就大家所見兒童和少年的行爲，他說的東西不無道理。
- 嘉穎，只因他有親身經驗。
- 駱君麗博士，只因她以實際的方式鋪陳理念。
- 我認爲嘉穎正談論的，自有他別出心裁的想法。他給你具體的方式以茲改進你的公開講演。如果你跟隨他所說的，我確信你將幹得不錯！：)
- 你大可與小馬的連結相配，它有關於怎樣克服懼怕公開講演的最有用的指點。[「小馬的連結」暗示駱博士]
- 我認爲嘉穎正確。事先排演並且多了解你的題目，確實是有幫助。在某程度上，我也同意莉莉，因為某些人是比其他人外向。但是作好準備和抱持良好態度，你就能夠有好的講演。完全避免它不是一個解決辦法!!
- 莉莉的想法描述人們不同的模式，因此她是我會相信的那個。

零分

代碼 0： 確認一位作者，但未有提供任何解釋。

- 心理學家 O.L.。
- 駱君麗博士。

回答含糊或不充分。

- 嘉穎，只因我同意他。
- 駱君麗博士，只因她是最好的。
- 嘉穎，只因他的意見有道理。

顯示對資料有不正確的理解，又或答案不合理、不切題。

- 心理學家 O.L.，只因他是我的心頭好。
- 嘉穎，原因是他教你怎樣避免公開演講。[不正確理解]
- 我會選小馬。[不是美琪提及的四位作者之一]

代碼 9： 沒有作答。

Differential Item Functioning as a Means to Examine the Strengths and Weaknesses of Electronic Reading Literacy of Macao Students

Pou-Seong Sit

Electronic reading assessment (ERA) is a new form of international assessment in the PISA 2009 Reading Literacy Study. Using students' online reading responses in the field trial, the present study examined Macao 15-year-old students' unexpected strengths and weaknesses through differential item functioning. Specifically, characteristics of ERA items identified as having country by item interactions were analyzed to throw light on the item format (multiple choice versus constructed response) and the reading aspects (i.e. access and retrieve, integrate and interpret, reflect and evaluate) that the students were good at and/or weak in. Through the use of released items after the field trial by the PISA Consortium, it is possible to examine item content so as to have a more concrete idea of Macao students' online Chinese reading capabilities and limitations.

Keywords: reading literacy, electronic assessment, differential item functioning, PISA

Pou-seong Sit, Assistant Professor, Faculty of Education, University of Macau

