

教育科學研究期刊 第六十二卷第三期

2017 年，62 (3)，1-23

doi:10.6209/JORIES.2017.62(3).01



從教師自編科學成就測驗之 Rasch 分析看教與學

陳映孜

國立中山大學
教育研究所

何曉琪

國立中山大學
人文創新與社會實踐研究中心

劉昆夏

國立中山大學
教育研究所

林煥祥

國立中山大學
通識教育中心

鄭英耀

國立中山大學
教育研究所

摘要

本研究採 Rasch 測量模式，分析教師自編科學成就測驗，以檢驗測驗試題的品質，並提供有用的評量回饋訊息，作為教師調整教學與修正評量之參考。研究對象為高雄市 8,936 位（男生 4,617 位、女生 4,319 位）七年級學生。研究工具為教師自編生物科成就測驗。測驗資料採試題反應理論的 Rasch 模式進行分析。結果發現：整體而言，教師自編科學成就測驗能區分出學生的能力程度。然而，有些試題的內容對女生較不公平。本研究也發現，教師對於部分試題難易度的認知與學生實際的作答反應存有若干差異。最後，本研究透過評量回饋圖表，包含學生能力與試題難度對照圖、學生作答診斷圖等，提出辨識學習基礎發展區和學習可能發展區的範圍之方式，作為教師規劃學生學習進程之參考。

關鍵詞：Rasch 測量、教師自編成就測驗、評量回饋、學習可能發展區、學習進程

壹、前言

課室評量 (classroom assessment) 維繫著教學品質之良窳。Black 和 Wiliam (2010) 回顧 250 篇有關課室評量的文獻，發現教師採用良好的評量實務，能有效促進學生的學習理解，其效果值 (effect size) 介於 .4 到 .7 之間，比起大多數的教學介入更能有效提升學生的學習成就。而一個好的課室評量，應充分連結學習目標與教學方法，以支持教師的教學和學生的學習 (Taylor & Nolen, 2008)。就整個教學歷程來看，從起點行為的診斷、教學目標的設定、到教學方法的採用，皆與課室評量息息相關。教師透過良好的評量方式，不僅可以瞭解學生已學會哪些知識技能 (assessment of learning)，進而可在學生學習的可能發展區 (the Zone of Proximal Development, ZPD) (Vygotsky, 1930-1935/1978)，選擇對學生有意義且能理解的學習內容，經由適當的教學引導與學習回饋，以促進學生的學習 (assessment for learning)，達到預定的學習目標。

就評量目的而言，一般可將評量分為著重在評定學生學習成果的總結性評量 (summative assessment)，以及著重學習診斷與教學改進功能的形成性評量 (formative assessment)。總結性評量雖具有評鑑教學計畫的實施成效、評量學生的學習成就、區分學生等第與篩選人才等功能，但卻無法提供足夠的訊息，作為學生精進學習與教師改進教學之參考。反之，形成性評量是適用於教與學過程中的評量，可即時確認學生的學習狀況，進而透過適當的協助，以促進學生的學習 (Bell & Cowie, 2001; Black, Harrison, Lee, Marshall, & Wiliam, 2004; Black & Wiliam, 1998)。在課堂中，若能將評量融入於教學過程，讓每個教學活動都伴隨著評量任務，不僅學生可透過評量任務的練習，促進對科學概念的理解 (Treagust, Jacobwitz, Gallagher, & Parker, 2001)，教師也可透過學生的反應，瞭解學生的學習情形，以決定何時教導學生新的概念與決定採用何種教學活動，以維持學生的學習動機和興趣 (Bell & Cowie, 2001)。

基本上，在整合評量於教學的過程中，教師對學生的評量方式、對評量結果的詮釋，以及對學生的回饋品質，皆攸關學生學習的成效 (Bell, 2007; Black et al., 2004)。可惜的是，在教師專業養成的階段，有關評量的相關課程並不受重視，導致教師的評量素養普遍不足 (Maclellan, 2004; Mertler, 2004)。一項於美國全國性的教師評量素養抽樣調查發現，在所有評量素養的向度中，教師在「與學生、家長或其他教育者溝通評量結果」的能力最低 (Plake & Impara, 1996)。換句話說，教師對於評量結果的解釋是有困難的，而這也直接影響到評量結果在後續的教學運用。

為了提升教師的評量實務能力，許多研究也紛紛著手開發教師評量素養培育課程。例如：Mertler (2009) 根據「學生教育評量的教師能力標準」(the standards for teacher competence in the educational assessment of students) 設計一套為期 2 週的在職教師課室評量工作坊，經實驗研究

發現，該工作坊課程能增進教師的評量知識與技能。Sato、Wei 和 Darling-Hammond (2008) 的 3 年縱貫研究也發現，教師透過參與全美專業教學標準委員會 (the national board for professional teaching standards) 的認證過程，可提升其評量實務的能力，包括採用各種評量的能力和運用評量訊息來支持學生的學習等。在國內，王子華、范雅晴和王國華 (2008) 運用「個人化」和「情境化」的網路評量和試題分析系統 (web-based assessment and test analysis) (Wang, Wang, Wang, Huang, & Chen, 2004)，發展的數理科在職教師評量素養培育模式，亦有助於提升教師的評量知能。而此模式運用於培育職前生物科教師的評量素養，也得到類似的成效 (Wang, Wang, & Huang, 2008)。

然而，目前的教師評量素養培育課程多著重在提升教師的評量知識，有關於如何在課堂中運用評量的回饋，以促進學生學習的技術，則較少有相關的文獻。就課堂中的形成性評量而言，評量結果的回饋乃扮演著促進學習的關鍵性角色 (Black et al., 2004)。良好的評量回饋不僅能讓學生知道自己的優缺點 (Black & Wiliam, 1998)，也可讓學生知道自己的表現是否達到預定的學習目標 (標準參照)、自己的表現和其他同學的比較 (常模參照)，以及自己的學習成長情形 (自我參照) (Bell, 2007)。不過，並非所有的回饋都可以促進學生的學習 (Hattie & Timperley, 2007; Kluger & DeNisi, 1996)。Kluger 和 DeNisi (1996) 的後設分析研究發現，雖然多數的回饋都有助於學習，且平均效果值達 .41，但仍有三分之一的回饋會降低學習表現。他們發現，當回饋只是單純的獎勵或評判，將降低回饋對學習的助益。反之，有助於學習的回饋方式，則多是將回饋聚焦於告訴學生如何達到預定目標的改進方法 (Lipnevich & Smith, 2009)。Hattie 和 Timperley (2007) 回顧了探討回饋效果的文獻，歸納出有效的回饋必須要能回答學生三個問題：一、學習的目標為何？二、要到達目標的途徑為何？三、目前應採取何種行動才能更接近目標？以 Vygotsky (1930-1935/1978) 的 ZPD 概念來說，良好的評量應該要能辨識出學生學習的 ZPD 範圍，亦即指出哪些知識概念是學生既有能力所應該學會的，哪些知識概念是學生在教師或同儕協助下，可以學會的，而評量回饋的目的就像是為學生搭建學習的鷹架，透過教師或同儕的協助，縮小學生目前已學會和未來預期學會的目標之差距，支持學生一步一步地朝學習目標邁進 (Shepard, 2005; Shute, 2008)。

可惜的是，在升學考試制度的影響下，教師很少運用課室評量的回饋技巧來幫助學生的學習。研究發現，教師為了配合國家考試的內容與形式，會傾向採用總結性成就測驗，著重於評量學生是否學會國家的課程目標，而較少運用形成性評量，提供學生學習改進方向 (Gioka, 2009; Ogan-Bekiroglu, 2009)。此外，教師因擔心自己的評量素養和學科專業知識不足 (Wang, Kao, & Lin, 2010)，以及編製測驗與實施評量的時間有限 (Graham, 2005)，往往採用傳統的教師自編成就測驗進行課室評量。以當前的臺灣教育現場為例，教師在課堂中所進行的評量，仍是習慣在單元教學後，依據課程綱要所訂定的能力指標和單元主題內容，自行編製評量測驗進行考試，並在考試結束後，採用測驗總分評定每位學生的成績，再逐題訂正

答案的對錯，以作為檢討學習成果的依據。然而，這種僅以考試總分回饋給學生的作法，對於改進學生的學習十分有限，且強調學生間的分數競爭，容易對學習低成就學生造成負面影響，導致習得無助感，甚至放棄學習（Black & Wiliam, 2010）。因此，在當前教育現場強調升學考試的氛圍，以及在評量實施時間有限的情況下，若能以目前教師習慣採用的自編成就測驗為評量工具，提供教師一套可兼顧升學考試內容與課堂學習診斷的評量分析方式，進而辨識學習基礎發展區（zone of fundamental learning development）和學習可能發展區（zone of proximal learning development）的知識概念範圍，或許是回應當前教育部精進教學政策，有關「提升學校紙筆測驗試題品質」、「進行學生學習評量之試題題型及測驗結果分析」和「建立教師教學評量回饋機制」（教育部，2008, p. 31624）的具體可行方式。過去研究曾以量化提示次數之方式（Budoff, 1987; Guthke, 1993）或以 Rasch 模式估算學習者達成學習任務的機率以表徵 ZPD 的範圍（Griffin, 2007; Warbruck & Stachelscheid, 2010），然前者提示次數實施不易，後者則缺乏實證證據支持。此外，Allal 與 Ducrey（2000）的研究提出兩種不同的 ZPD 解釋觀點，其一認為學習者的 ZPD 可被視為是可測量的個人特點，並在教學中保持穩定性；其二則是將 ZPD 視為互動式形成性評量，使學習者在教學環境中保有持續性的師生互動。本研究試圖在實際教學過程提出 ZPD 具體可行的操作方法，以作為教師教學與學生學習診斷的參考。

近年來，美國加州大學柏克萊分校的評估與評量研究中心（the Berkeley Evaluation and Assessment Research Center, BEAR Center）運用試題反應理論的 Rasch 模式（Rasch, 1960），已發展出一套可促進學習的評量系統（Wilson & Scalise, 2006）和免費的試題分析套裝軟體（Kennedy, Wilson, & Draney, 2008）。此評量系統兼具形成性評量和總結性評量的功能，可進行學生的學習診斷和檢核學生的學習成效（Wilson, 2008），且已陸續被運用在數學（Wilson & Carstensen, 2005）、化學（Claesgens, Scalise, Wilson, & Stacy, 2009）、地球科學（Briggs, Alonzo, Schwab, & Wilson, 2006）和物理（Alonzo & Steedle, 2009）等領域的評量。此外，由於 Rasch 測量模式具有客觀、等距的測量特性，特別適用於評量具有階層性的科學知識概念，近年來已逐漸被應用於分析和發展科學評量工具（e.g., Boone & Scantlebury, 2006; Boone, Townsend, & Staver, 2011），以及評量科學概念的學習進程（learning progressions）（e.g., Lee & Liu, 2010），且可融入於一般的課堂教學中，提供即時的評量回饋圖表，幫助教師進行教學與評量的設計、實施、反思與修正（Wilson et al., 2010）。Fulmer（2014）即結合 Rasch 測量模式與學習進程的概念，以新加坡 174 位中學生為對象，驗證力學概念工具（Force Concept Inventory, FCI）對於學生力與運動概念上學習進程的改變。基於此，本研究目的旨在運用 Rasch 測量模式，分析學校教師自編的成就測驗，並從學生實際的作答反應，檢核教師的命題品質，提供學生的學習診斷圖和辨識學習基礎發展區與學習可能發展區的範圍，以作為教師修正評量、調整教學與規劃學生的學習進程之參考。

學習進程是學生「從對概念的粗淺認識到精熟理解」之學習成長歷程的一種描述 (Bertenthal & Wilson, 2006)。其主要特徵包括：一、著重在描述一些基礎的且具有發展性規律的概念；二、根據學生在某一段期間的學習歷程，設定理解程度的最高層次（該期間學習的最終目標）和最低層次（該期間學習初期所需具備的先備知識）；三、根據學生對概念的理解程度，在最高層次和最低層次的理解之間區分不同的層次；四、隨著教學的介入，學生的學習進程會改變，且學習進程的發展並非只有單一軌跡 (Duncan & Hmelo-Silver, 2009)。其中，本研究根據學生答對試題的機率，提出學習基礎發展區（學生基本／最低的理解層次）和學習可能發展區（學生透過學習可達到最高理解層次）之範圍的辨識方式，在教學現場將學習進程概念轉化為可具體操作的評量、教學與回饋機制，也可為當前補救教學找出具體可行的步驟。事實上，關於國際間對於學校教育中科學概念的學習進程研究已日趨具體明確，如 Plummer 等 (2015) 發展太陽系如何形成 (Formation of the Solar System) 的學習進程，Fulmer (2014) 和 Neumann、Fulmer 與 Liang (2013) 以學習進程為架構，針對力與運動編製量表，Testa、Galano、Leccia 與 Puddu (2015) 發展四季變化、日蝕、月蝕與月亮週期 (change of seasons, solar and lunar eclipses, and moon phase) 的學習進程，Furtak、Morrison 與 Kroog (2014) 發展自然選擇 (natural selection) 的學習進程等，足見此概念在教學、評量上的重要性，然而臺灣目前未有相關研究產出，本研究透過嚴謹創新的研究設計，試圖結合學習進程與 Rasch 測量模式之研究，且研究成果尚屬符合預期，將有助於科學領域中學習進程的建構與相關評量發展的依據。

貳、Rasch 測量模式

Rasch 模式 (Rasch, 1960) 是丹麥的數學家 G. Rasch 所提出的測量模式。此模式是以機率的方式來表徵受測者的作答反應。以非對即錯的二分題（只有二種計分方式）為例，在一個理想的測量情境下，Rasch 主張考生 n 在題目 i 答對（計為 1 分）或答錯（計為 0 分）的機率，主要受考生的能力 θ_n 和題目的難度 δ_i 二個因素影響，並以 P_{ni1} 和 P_{ni0} 分別表示考生 n 在題目 i 上得 1 分和 0 分的機率，其數學式可以 log 函數表示為：

$$\log(P_{ni1} / P_{ni0}) = \theta_n - \delta_i \quad (1)$$

由於 $P_{ni1} + P_{ni0} = 1$ ，根據公式(1)可推導出：

$$P_{ni1} = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)} \quad (2)$$

其中 $\exp$ 是自然指數 2.718， $\exp(\)$ 表示自然指數的次方。從公式(2)的數學式可發現，當考生 n 的能力 θ_n 和題目 i 的難度 δ_i 已知時，帶入公式(2)即可算出考生 n 作答題目 i 的答對機率，其

關係可以圖 1 表示。

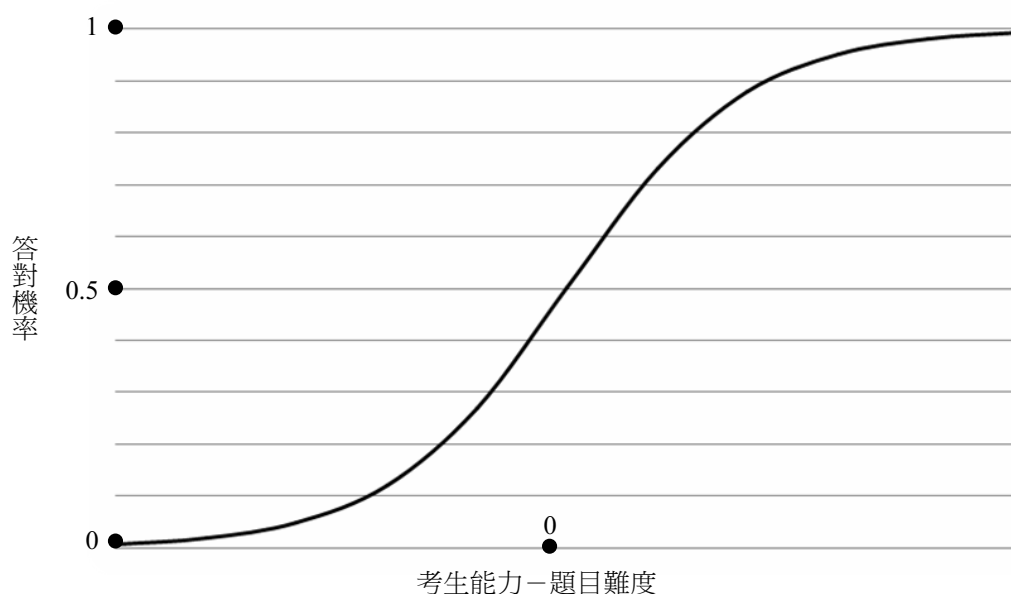


圖1. 考生能力 θ_n 減題目難度 δ_i 與答對機率的關係

從圖 1 可看出，當 θ_n 大於 δ_i 時（能力高於難度），考生 n 在題目 i 上答對的機率會高於 .5；當 θ_n 小於 δ_i 時（能力低於難度），考生 n 在題目 i 上答對的機率則低於 .5；當 θ_n 等於 δ_i 時（能力和難度相當），考生 n 在題目 i 上答對的機率則剛好等於 .5。

參、研究方法

一、研究對象

本研究的對象為高雄市 10 個行政區，20 所國中，279 個班級的七年級學生，共計 8,936 位（男生 4,617 位、女生 4,319 位）。所有的學生於 96 學年度下學期期末考後 1 週，由各校教師安排統一時間進行「生物科成就測驗」的施測。

二、研究工具

本研究所採用的「生物科成就測驗」係由高雄市國民教育輔導團所組成的國中生物教師團隊，依據自然與生活科技領域課程綱要中有關生物科之內容，並參考國中基本學測試題和各國中歷年試題所編製。試題內容包含「生殖、遺傳與演化」、「生命的多樣性」、「生物

與環境」、「動物的構造與功能」和「植物的構造與功能」等五個單元主題。試題的「認知向度」包含「記憶」、「理解」、「應用」和「分析」，每一試題在「單元主題」與「認知向度」的雙向細目分析如表 1 所示。此外，為了讓整份測驗能兼具不同難度的題目，編題教師根據自己的教學經驗，將試題區分為「難」、「中」、「易」三種程度，試題的難度分配如表 2 所示。整份測驗共有 34 題四選一的選擇題，經以高雄市 322 位國中學生進行預試，試題難度適中（ $.24 < P < .74$ ），且可鑑別高、低分組學生的能力（ $t=39.37, p < .001$ ）。

表 1

生物試題在「單元主題」與「認知向度」之雙向細目分析

單元主題	認知向度			
	記憶	理解	應用	分析
生殖、遺傳與演化	2	1, 17, 22	20	19, 21
生命的多樣性		3, 5	4, 6	
生物與環境	7, 24	8, 9, 10	11	23
動物的構造與功能	15, 27	12, 13, 25, 31, 34	14, 29, 30	18
植物的構造與功能	26	28, 33	16	32

註：數字代表試題的編號。

表 2

教師對生物試題的難度預期

單元主題	難度		
	易	中	難
生殖、遺傳與演化	2, 21	1, 17, 20, 22	19
生命的多樣性		3, 5	4, 6
生物與環境	7, 24	8, 10, 11, 23	9
動物的構造與功能	14, 15, 27, 34	12, 13, 18, 25, 31	29, 30
植物的構造與功能	16	26, 28, 33	32

註：數字代表試題的編號。

三、資料分析

本研究的測驗資料採二元計分，答對計 1 分，答錯計 0 分，因此採用 Rasch 模式（Rasch, 1960）進行學生作答反應的分析。模式－資料適配度（model-data fit）分析和性別的差異試題功能（differential item functioning, DIF）（Holland & Wainer, 1993）分析，採用套裝軟體 ConQuest（Wu, Adams, & Wilson, 2007）進行。

模式－資料適配度分析的主要目的，在檢驗本研究的「生物科成就測驗」中的題目是否符合 Rasch 模式的單向度假設。當學生的作答反應資料能符合單向度假設，表示這些題目背後所測量到的是相同的構念，亦即生物科的能力，因此才可進行試題難度的比較，以及學生能力和試題難度對照圖分析。適配度指標採用 infit (weighted) 及 outfit (unweighted) 均方誤 (mean square error, MNSQ) 統計量，其數值介於 0.7-1.3 之間，表示模式－資料適配度良好 (Linacre & Wright, 1994)。

性別的 DIF 分析之目的，在檢驗不同的試題對於男、女生是否具有公平性。換句話說，DIF 分析可檢驗試題的難度對於男、女生是否有超出預期的差異。所謂超出預期的差異是指，DIF 分析在比較某一試題對男、女生的難度差異時，是建立在男、女生具有相等的能力之前提下。也就是說，某一試題對男、女生的難度差異，已排除男、女生既有的能力之影響，如此可避免因取樣的偏誤所造成的估計誤差。就一般的 DIF 分析而言，當兩個群體在某個試題的難度估計值差距達 0.5 logits 時，表示此試題已具有一定程度的 DIF (Wang, 2008)。此外，本研究採用 BEAR Center 所開發的免費套裝軟體 ConstructMap (Kennedy et al., 2008) 進行學生的作答診斷分析，進而辨識學生的學習基礎發展區與學習可能發展區的範圍，以作為教師實施補救教學與規劃學生的學習進程之參考。

肆、結果與討論

一、模式－資料適配度

表 3 是每一試題的難度估計值與適配度。由表 3 可知，整體而言，所有題目的適配度佳，infit MNSQ 介於 0.83~1.17 之間，outfit MNSQ 介於 0.76~1.21 之間。表示所有的試題符合 Rasch 模式單向度的假設，適合用來測量生物科的知識概念、進行試題難度的比較，以及學生能力和試題難度的對照分析。另外，針對 34 題信度分析的結果為 .85。

二、性別的差異試題功能

表 3 的「難度差異」欄位表示，每一試題對於男生的難度估計值減對於女生的估計值。當此差值為正值時，表示該題目對於男生而言較難；負值時，表示對女生而言較難。由表 3 可知，第 18 題的難度差異值為-0.65，表示此試題對女生而言明顯偏難。分析男、女學生在第 18 題的期望得分發現（如圖 2），對於相同能力水準的男、女學生而言，男生在第 18 題的期望得分都比女生高，顯示第 18 題較有利於男生作答。進一步檢視第 18 題的試題內容（見圖 3）可發現，由於此題需要由圖（二）的染色體數目的改變來判斷人體的哪一個構造在進行此項工作，相較於男生而言，青春期的國中女生，在直視圖（一）的男性生殖系統時，可能會產

表 3

整體試題難度估計值、適配度與男、女生的試題難度估計值之差異

試題	難度 (logits)	infit MNSQ	outfit MNSQ	難度差異 (男生－女生)
I1	-1.53	0.92	0.77	0.05
I2	-1.04	0.90	0.79	0.34
I3	-0.78	1.05	1.10	0.14
I4	0.50	1.07	1.10	0.04
I5	1.10	1.03	1.11	-0.08
I6	-0.97	1.00	1.04	0.00
I7	0.00	0.94	0.91	-0.08
I8	0.58	1.17	1.21	0.10
I9	-0.46	0.94	0.89	0.07
I10	-0.94	0.92	0.89	-0.01
I11	-0.47	1.05	1.09	-0.16
I12	-0.04	1.01	1.01	0.13
I13	1.11	0.97	1.06	-0.32
I14	0.00	1.03	1.03	-0.08
I15	0.36	1.16	1.21	-0.21
I16	-1.05	0.93	0.89	0.34
I17	-0.07	0.83	0.76	0.00
I18	0.08	0.94	0.92	-0.65
I19	0.14	1.05	1.06	0.17
I20	0.23	0.92	0.89	-0.09
I21	0.16	1.02	1.03	-0.10
I22	0.59	0.97	0.99	-0.25
I23	-0.54	0.90	0.87	0.13
I24	-0.81	0.91	0.88	0.25
I25	-0.79	0.86	0.77	-0.00
I26	0.99	1.08	1.16	0.01
I27	0.68	1.11	1.16	-0.12
I28	1.08	1.03	1.08	-0.24
I29	-0.14	0.98	0.97	0.10
I30	0.85	1.14	1.20	0.33
I31	0.74	1.05	1.10	-0.09
I32	0.44	1.05	1.06	0.00
I33	-0.45	0.91	0.86	0.25
I34	0.44	1.03	1.04	0.00

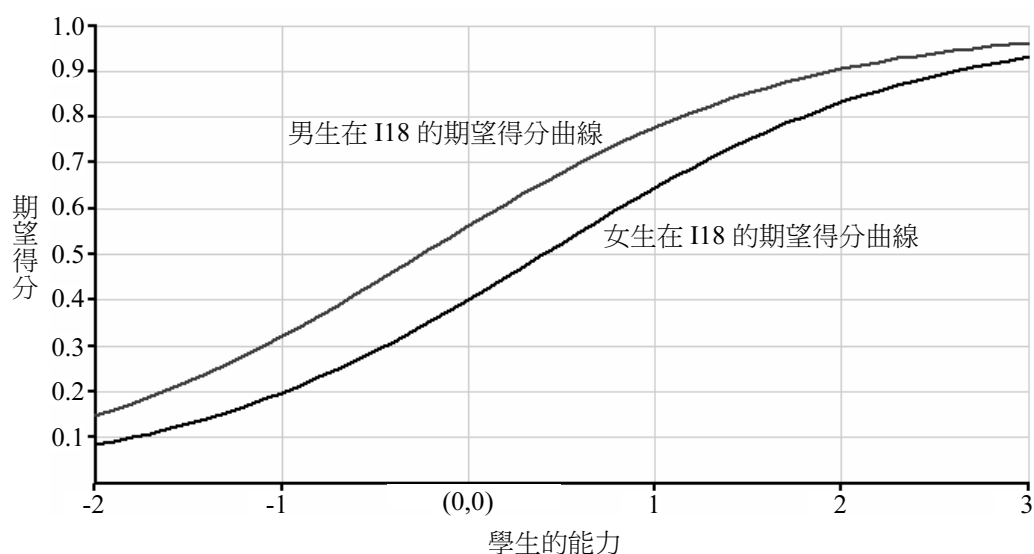
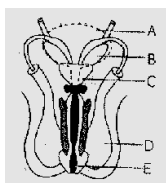


圖2. 男生和女生在第18題的期望得分比較

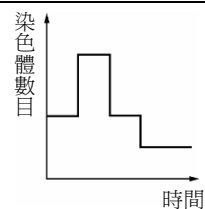
I18.圖（一）為男性的生殖系統，圖（二）為細胞進行分裂過程中染色體數量的變化圖。

圖（一）中哪一部分能進行如圖（二）所示的分裂方式？

(A)A (B)B (C)C (D)D



圖（一）



圖（二）

圖3. 第18題試題內容

生一些心理上的障礙，以致於干擾其作答，因而較不容易答對此題目。因此，根據此評量回饋訊息，教師未來在命題時，應特別留意試題的內容是否會對不同群體的學生造成不公平。

三、學生能力與試題難度對照圖

圖 4 是學生能力與試題難度分布對照圖。圖左邊的“X”表示學生的生物科能力之分布，分布在愈上面的學生，表示其生物科能力愈高；圖右邊是每一試題的難度分布，相同單元主題的題目排在同一直欄，分布在愈上面的題目，表示試題難度愈難。圖中間垂直的虛線表示學生能力和試題難度的共同單位量尺，以 logit 表示。由表 3 和圖 4 可看出，所有的試題難度介

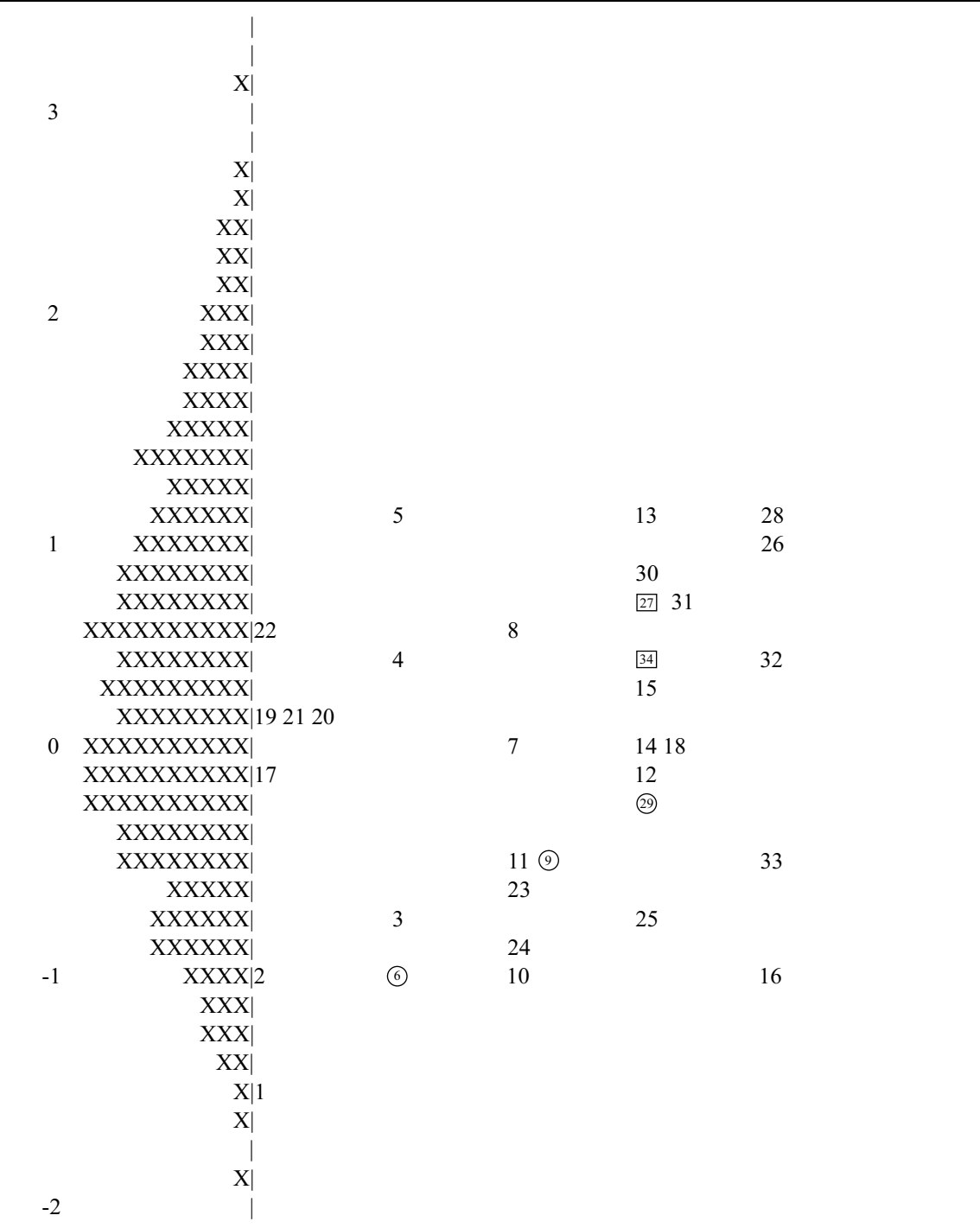


圖4. 學生能力與試題難度分布對照圖。圖左邊每一個“X”表示49.5位學生；圖右邊的數字表示每一試題的編號。○內的試題表示教師認為困難但學生實際作答結果卻偏簡單的題目；□內的試題表示教師認為簡單但學生實際作答結果卻偏困難的題目。

於-1.53 logits (第 1 題)~1.11 logits (第 13 題)之間 ($M=0$, $SD=0.72$)，學生的生物科能力分布大約介於-2 logits~3 logits 之間 ($M=0.39$, $SD=1.00$)。整體而言，所有試題的平均難度略低於所有學生的平均能力，試題難度的分布大致吻合學生能力的分布，適合用來評量不同能力的學生。就每一個單元主題而言，「生殖、遺傳與演化」最容易的題目為第 1 題 (-1.53 logits)，約有 84%的學生答對，最難的為第 22 題 (0.59 logits)，約有 46%的學生答對；「生命的多樣性」最容易的題目為第 6 題 (-0.97 logits)，約有 76%的學生答對，最難的為第 5 題 (1.10 logits)，約有 36%的學生答對；「生物與環境」最容易的題目為第 10 題 (-0.94 logits)，約有 75%的學生答對，最難的為第 8 題 (0.58 logits)，約有 46%的學生答對；「動物的構造與功能」最容易的題目為第 25 題 (-0.79 logits)，約有 73%的學生答對，最難的為第 13 題 (1.11 logits)，約有 35%的學生答對；「植物的構造與功能」最容易的題目為第 16 題 (-1.05 logits)，約有 77%的學生答對，最難的為第 28 題 (1.08 logits)，約有 36%的學生答對。就試題內容而言，以「生殖、遺傳與演化」單元主題為例，試題由易到難依序為第 1 題 (演化)、第 2 題 (演化)、第 17 題 (生殖)、第 19 題 (生殖)、第 21 題 (遺傳)、第 20 題 (生殖)、第 22 題 (遺傳)。由此可知，對國中學生而言，「演化」的試題較容易，其次為「生殖」，而「遺傳」的試題較難。進一步檢視試題內容發現，「演化」主要考的概念為適者生存的「天擇說」，例如蛾的顏色改變；「生殖」考的概念為細胞的分裂；「遺傳」考的是染色體與基因的概念。就概念的層次來看，從「蛾」到「細胞」再到「染色體」與「基因」，隱含著從巨觀概念到微觀概念，從具體概念到抽象概念。Duncan 和 Reiser (2007) 的研究發現，基因概念對學生而言是一個不容易學習的概念。而此概念難以理解的原因主要來自學生不瞭解基因的功能和基因所組成的結構。其中結構從微觀到巨觀依序為基因、蛋白質、細胞、組織、器官、有機體，愈是微觀的概念，學生愈不容易理解，此結果也間接支持了本研究的發現。

四、試題的預期難度和實際難度比較

本研究進一步比較圖 4 學生實際的作答反應所估計出來的試題難度排序和表 2 教師對於試題難度的預期可發現，原本被教師認為「難」的第 6 題、第 9 題和第 29 題，學生的實際作答結果卻偏簡單，答對率分別為 76%、67%和 61%；原本被教師認為「易」的第 27 題和第 34 題，學生的實際作答結果卻偏難，答對率分別為 44%和 49%。此結果顯示，教師對學生的知識概念學習的認知與學生實際的表現有不一致的情形。Otero 和 Nathan (2008) 的研究發現，職前科學教師對於學生既有的科學先備知識所知有限。換句話說，職前科學教師並不清楚學生已學會哪些知識概念或是存有哪些科學迷思概念。究其原因可能是因教師的教學經驗不足，與學生的互動機會少，因此不瞭解學生學習科學概念的方式。以第 27 題為例 (如圖 5)，教師認為此題目屬於記憶題，因此只要教學生背誦即可。但對學生而言，要答對此題目除了要知道各種血球的功能外，還需要知道各種血球微觀的形狀，只要其中一項不清楚，就無法

I27.如圖為人類血球模式圖，下列敘述何者正確？

- (A)甲能運送氧氣
- (B)乙能產生抗體
- (C)丙具有防禦疾病的功能
- (D)丁能從微血管壁擠出吞噬細菌

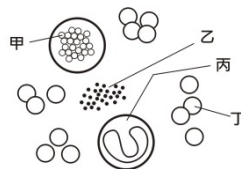


圖5. 第27題試題內容

正確的作答。顯然，教師並無法完全掌握學生對於知識概念的學習情形。而透過本研究的試題難度排序分析，正可幫助教師找出問題點，進而調整教學方式，以促進學生的學習理解。

五、學生作答診斷分析

圖 6 是學生 s26 在生物試題的作答反應之診斷圖。圖 6 的縱軸表示學生能力與試題難度的共同單位量尺 (logit)，圖形中的數字為試題編號，愈下面表示試題愈簡單，愈上面表示試題愈難。圖形中間的“XXX”表示 s26 的能力位置 (0.06 logits)，兩水平點線的區間表示 s26 能力估計值 ± 0.5 logits 的範圍。根據公式(2)的計算，當學生的能力值 θ_n 等於某一試題的難度值 δ_i 時，學生在該題答對的機率會等於 .50；當 $\theta_n - \delta_i = 0.5$ 時，學生在該題答對的機率約等於 .62；當 $\theta_n - \delta_i = -0.5$ 時，學生在該題答對的機率約等於 .38。以圖 6 來說，當 s26 作答兩水平點線區間的題目時，例如第 18 題，由於題目的難度和 s26 的能力差不多，因此，理論上 s26 答對的機率大約 50%；當 s26 作答水平點線上方的題目，例如第 8 題，由於題目的難度比 s26 的能力高 0.5 logits 以上，因此，理論上 s26 答對的機率應該會低於 .38，換句話說，s26 較可能答錯第 8 題。當 s26 作答水平點線下方的題目，例如第 23 題，由於題目的難度比 s26 的能力低 0.5 logits 以上，因此，理論上 s26 答對的機率應該會高於 .62，換句話說，s26 較可能答對第 23 題。然而，實際上 s26 的作答反應並不完全符合理論預期。以圖 6 中間垂直虛線將實際答對和答錯的題目區分為兩部分，虛線左邊的區域為實際答對的題目，虛線右邊的區域為實際答錯的題目。透過水平點線和垂直虛線的劃分，除了介於兩水平點線區間的題目外，其他題目可區分為四個象限，左下角表示，題目難度比學生能力低，且學生答對的題目，可稱為「應對也對」區（包括第 3、24、25、2、16 題）；右上角，表示題目難度比學生能力高，且學生答錯的題目，可稱為「應錯也錯」區（包括第 8、22、27、31、26、28 題）。因此，若是題目落在左下角和右上角，表示學生的作答反應吻合 Rasch 理論模式的預期；而左上角表示題目難度比學生能力高，但學生卻答對的題目，可稱為「應錯卻對」區（包括第 30、5、13 題），其原因可能是學生猜測或背誦答案等；右下角表示題目難度比學生能力低，但學生卻答錯的題目，可稱為「應對卻錯」區（包括第 23、6、10、1 題），其原因可能是學生粗心、無意願作答或時間不夠等。因此，若是題目落在左上角和右下角，則表示學生的作答反應不吻合

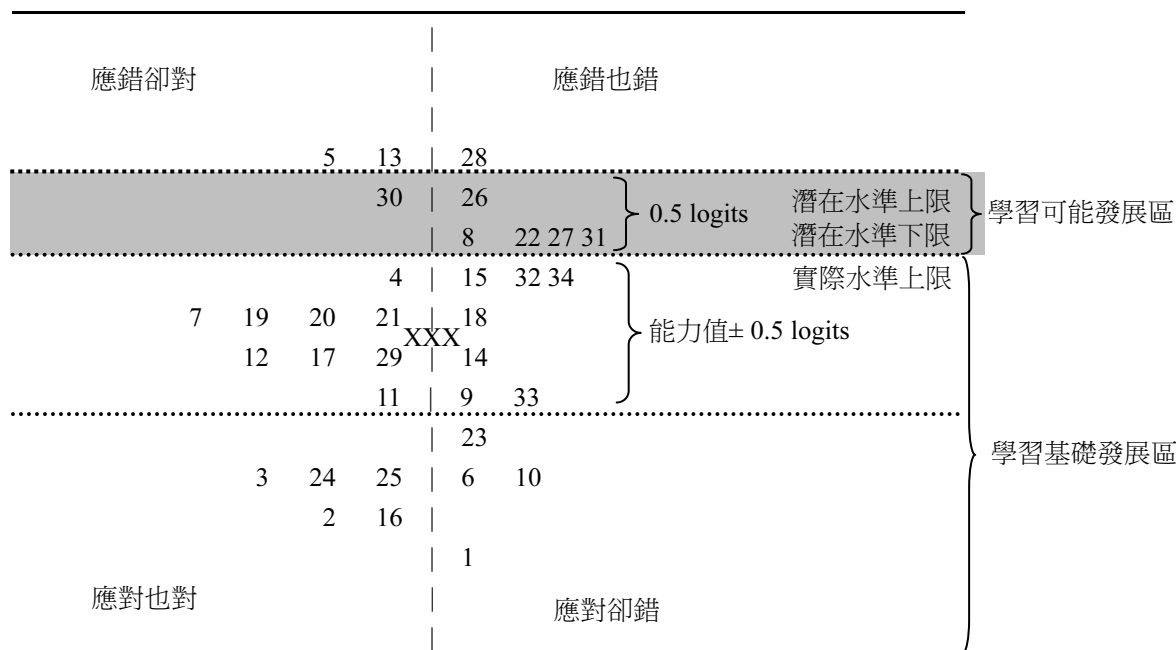


圖6. 學生s26的作答診斷圖。每行間距0.255 logits。

Rasch 理論模式的預期，教師可根據此訊息進一步檢視試題內容和訪談學生，以釐清學生是否真正理解該題目的知識概念，進而實施適性的補救教學。

六、學習基礎發展區和學習可能發展區

根據 Vygotsky (1930-1935/1978) 的 ZPD 定義，ZPD 是指介於學生獨自表現的實際水準 (actual level) 和在他人的協助下所表現的潛在水準 (potential level) 之間的距離。然而，ZPD 的概念至今仍僅處於文字的抽象描述，尚未具體落實於教學實務中。為了突破此困境，本研究基於 Rasch 分析可將學生能力和試題難度置於相同單位量尺之特性，試圖將 ZPD 的抽象概念予以具體化，如圖 6 灰底處。就試題的作答而言，學生獨自表現的實際水準，可視為是學生在既有的能力水準下應該答對的題目。以圖 6 的學生 s26 為例，“XXX”的位置可表示為 s26 的實際水準中心，上方的水平點線位置 (s26 能力+0.5 logits) 可視為 s26 的實際水準上限，當試題的難度高於 s26 的實際水準上限時，例如第 8 題，理論上 s26 的答對機率會低於 .38，因此，以 s26 既有的能力是不容易答對的。不過，經過教師或同儕的指導後，s26 可能就能理解第 8 題的知識概念，因此可視為在他人的協助下所表現的潛在水準。以圖 6 為例，s26 實際水準上限的水平點線上方，可視為 s26 的潛在水準下限，基本上，在潛在水準下限附近之試題的知識概念，雖然對於 s26 而言是有點難，不過，經過他人的教導後，s26 應該是可以理解的。然而，隨著試題的難度往上增加，知識概念就愈難理解，當潛在水準下限再往上+0.5 logit 至水平虛線“---”的位置時，此時，水平虛線以上的題目難度已高出 s26 能力 1.0 logits 以上，例

如第 28 題，根據公式(2)的計算，s26 答對此題的機率將低於 .27，表示此題目的知識概念對 s26 而言確實太難，就算有教師或同儕的指導，可能也不容易理解。因此，潛在水準下限往上 +0.5 logit 的水平虛線之位置可視為潛在水準的上限，而介於潛在水準上、下限之間的區域，即可視為 s26 的學習可能發展區，而在學習可能發展區以下的區域，則可視為學習基礎發展區。因此，教師可根據此兩個區域內的試題概念之難度階層，依序由簡單的概念到困難的概念，規劃學生的學習進程，進行適性的教學。

伍、結論與建議

本研究以 Rasch 模式分析學校教師自編的生物科成就測驗，結果發現教師所編製的試題符合 Rasch 模式單向度的假設，表示所有的題目都可測量到相同的能力，亦即生物科能力。此結果顯示，教師對於生物科的內容知識具有一定程度的理解，因此，教師所編製的試題都能測量到所欲測量的生物科能力。此外，所有試題的難度分布大致也和學生的能力分布相呼應，表示教師自編的測驗，適用於評量不同能力程度的學生。然而，雖然教師能編製出不同難度的試題，但教師對於試題難、易度的預期與學生實際的作答反應卻存有若干差異。例如，教師認為記憶層次的題目是簡單的，但學生的實際表現卻不如教師預期。因此，教師在教導學生這些知識概念時，仍需注重學生的理解，方能有助於學生將知識概念儲存於長期記憶中。而性別的 DIF 分析亦發現，有些試題的內容對女學生的作答較不公平，此訊息亦可作為教師未來編製或修改試題之參考。

再者，本研究也發現「生殖、遺傳與演化」的試題內容之難度排序，由易到難依序為「演化」、「生殖」、「遺傳」。其背後隱含著愈是巨觀的概念，學生愈容易理解；愈是微觀的概念，學生愈難理解。基於此，教師進行教學時，可從巨觀的概念先教，再依序進入微觀的概念，以幫助學生逐步理解概念。此外，本研究透過 Rasch 模式分析所提供的學生能力和試題難度分布對照圖，除了可比較學生間的能力差異外，亦可藉由學生能力和試題難度的比較，算出每位學生答對每一個題目的機率，進而瞭解學生對於每一個知識概念的精熟程度，兼具評量結果解釋的常模參照和標準參照的功能。再者，作答診斷圖可針對每一位學生的評量結果進行個別診斷，提供「應對也對」、「應錯也錯」、「應錯卻對」與「應對卻錯」四個面向的分析，可幫助教師釐清每一位學生真正理解的概念有哪些，需要進行補救教學的概念有哪些，以及可延伸學習的概念有哪些，進而實施適性的教學。簡言之，本研究根據學生答對試題的機率，提出學習基礎發展區和學習可能發展區之範圍的辨識方式，讓長期處於理念層次的學習可能發展區之概念，能成為教學上可操作的實務層次，一方面可作為教師進行科學教學時，依據科學概念發展層次建構教學模組之參考實施階段性教學的參考，另一方面也可透過實際的教師或同儕指導，以檢驗學生在學習可能發展區的學習成效，進而提供班級內實施適性化

之學習輔導與補救。未來也可透過對學生學習進程的具體描述，進一步建構學生學習的階段性和次序性，以作為設計課程、教學和評量之依據，進而促進課程、教學和評量之間的連結。

值得注意的是，本研究構想「編制有效評量學生學習表現、提供及時回饋機制並作為教師教學調整基礎」與近 10 年來科學教育領域中大力提倡的學習進程（Learning Progressions）是息息相關、相互呼應的。學習進程的概念源自於 2006 年美國國家研究委員會（National Research Council, NRC）出版的 *Systems for State Science Assessments* 和 2007 出版的 *Taking Science to School: Learning and Teaching Science in Grades K-8* 報告書中，先後建議美國州政府應透過學生的學習進程，將課程、教學和評量做有效的連結（Shouse, Schweingruber, & Duschl, 2007），可見學習進程對教學實務之重要性。再者，2013 年美國所公布的 The Next Generation Science Standards 已將學習進程視為科學核心概念學習所應依循的路徑。而國際學術期刊 *Journal of Research in Science Teaching* 與 *Science Education* 分別於 2009 年與 2015 年以學習進程為核心刊登相關研究（e.g., Duncan & Gotwals, 2015; Hammer & Sikorski, 2015; Mohan, Chen, & Anderson, 2009; Songer, Kelcey, & Gotwals, 2009; Steedle & Shavelson, 2009）。

參考文獻

一、中文文獻

王子華、范雅晴、王國華（2008）。數理科在職教師評量素養培育模式之探究。《科學教育學刊》，16（1），25-51。doi:10.6173/CJSE.2008.1601.04

【Wang, T.-H., Fan, Y.-C., & Wang, K.-H. (2008). Implementation of web-based assessment system in improving in-service teachers' assessment literacy. *Chinese Journal of Science Education*, 16(1), 25-51 doi:10.6173/CJSE.2008.1601.04】

教育部（2008）。教育部補助辦理精進教學要點修正規定。《行政院公報》，014（215）。取自 https://gazette.nat.gov.tw/EG_FileManager/eguploadpub/eg014215/ch05/type2/gov40/num12/Eg.pdf

【Ministry of Education. (2008). Revised regulations governing Ministry of Education subsidies for instruction improvement. *The Executive Yuan Gazette Online*, 014(215). Retrieved from https://gazette.nat.gov.tw/EG_FileManager/eguploadpub/eg014215/ch05/type2/gov40/num12/Eg.pdf】

二、外文文獻

Alonzo, A., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education*, 93(3), 389-421. doi:10.1002/sce.20303

Allal, L., & Ducrey, G. P. (2000). Assessment of-or in-the zone of proximal development. *Learning and Instruction*, 10(2), 137-152. doi:10.1016/S0959-4752(99)00025-0

Bell, B. (2007). Classroom assessment of science learning. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 965-1006). Mahwah, NJ: LEA.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85(5), 536-553. doi:10.1002/sce.1022

Bertenthal, M. W., & Wilson, M. R. (2006). *Systems for state science assessment*. Washington, DC: National Academy Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 8-21. doi:10.1177/003172170408600105

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74. doi:10.1080/0969595980050102

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81-90. doi:10.1177/003172171009200119

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science

- education research utilizing multiple-choice tests. *Science Education*, 90(2), 253-269. doi:10.1002/sce.20106
- Boone, W. J., Townsend, J. S., & Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: An exemplar utilizing STEBI self-efficacy data. *Science Education*, 95(2), 258-280. doi:10.1002/sce.20413
- Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, 11(1), 33-63. doi:10.1207/s15326977ea1101_2
- Budoff, M. (1987). The validity of learning potential assessment. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 53-81). New York, NY: Guilford Press.
- Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry: The perspectives on chemists. *Science Education*, 93(1), 56-85. doi:10.1002/sce.20292
- Duncan, R. G., & Gotwals, A. W. (2015). A tale of two progressions: On the benefits of careful comparisons. *Science Education*, 99(3), 410-416. doi:10.1002/sce.21167
- Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal for Research in Science Teaching*, 46(6), 606-609. doi:10.1002/tea.20316
- Duncan, R. G., & Reiser, B. J. (2007). Reasoning across ontologically distinct levels: Students' understandings of molecular genetics. *Journal of Research in Science Teaching*, 44(7), 938-959. doi:10.1002/tea.20186
- Fulmer, G. W. (2014). Validating proposed learning progressions on force and motion using the force concept inventory: Findings from Singapore secondary schools. *International Journal of Science and Mathematics Education*, 13(6), 1235-1254. doi:10.1007/s10763-014-9553-x
- Furtak, E. M., Morrison, D., & Kroog, H. (2014). Investigating the link between learning progressions and classroom assessment. *Science Education*, 98(4), 640-673. doi:10.1002/sce.21122
- Gioka, O. (2009). Teacher or examiner? The tensions between formative and summative assessment in the case of science coursework. *Research in Science Education*, 39(4), 411-428. doi:10.1007/s11165-008-9086-9
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21(6), 607-621. doi:10.1016/j.tate.2005.05.001

- Griffin, P. (2007). The comfort of competence and the uncertainty of assessment. *Studies in Educational Evaluation*, 33(1), 87-99. doi:10.1016/j.stueduc.2007.01.007
- Guthke, J. (1993). Developments in learning potential assessment. In J. H. M. Hamers & K. Sijtsma (Eds.), *Learning potential assessment: Theoretical, methodological and practical issues* (pp. 43-67). Amsterdam, The Netherlands: Swets & Zeitlinger.
- Hammer, D., & Sikorski, T. R. (2015). Implications of complexity for research on learning progressions. *Science Education*, 99(3), 424-431. doi:10.1002/sce.21165
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. doi:10.3102/003465430298487
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kennedy, C. A., Wilson, M., & Draney, K. (2008). *ConstructMap* [Computer Program]. UC Berkeley, CA: BEAR Center.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284. doi:10.1037/0033-2909.119.2.254
- Lee, H.-S., & Liu, O. L. (2010). Assessing learning progression of energy concepts across middle school graded: The knowledge integration perspective. *Science Education*, 94(4), 665-688. doi:10.1002/sce.20382
- Linacre, J. M., & Wright, B. D. (1994). Chi-square fit statistics. *Rasch Measurement Transactions*, 8(2), 350.
- Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, 15(4), 319-333. doi:10.1037/a0017841
- Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education*, 20(5), 523-535. doi:10.1016/j.tate.2004.04.008
- Mertler, C. A. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(1), 49-64.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101-113. doi:10.1177/1365480209105575
- Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*, 46(6),

- 675-698. doi:10.1002/tea.20314
- Neumann, I., Fulmer, G. W., & Liang, L. L. (2013). Analyzing the FCI based on a force and motion learning progression. *Science Education Review Letters*, 2013, 8-14. doi:10.18452/8208
- Ogan-Bekiroglu, F. (2009). Assessing assessment: Examination of pre-service physics teachers' attitudes towards assessment and factors affecting their attitudes. *International Journal of Science Education*, 31(1), 1-39. doi:10.1080/09500690701630448
- Otero, V. K., & Nathan, M. J. (2008). Preservice elementary teachers' views of their students' prior knowledge of science. *Journal of Research in Science Teaching*, 45(4), 497-523. doi:10.1002/tea.20229
- Plake, B. S., & Impara, J. C. (1996). Teacher assessment literacy: What do teachers know about assessment? In G. Phye (Ed.), *Handbook of classroom assessment* (pp. 53-68). San Diego, CA: Academic Press. doi:10.1016/B978-012554155-8/50005-3
- Plummer, J. D., Palma, C., Flarend, A., Rubin, K., Ong, Y. S., Botzer, B., ...Furman, T. (2015). Development of a learning progression for the formation of the solar system. *International Journal of Science Education*, 37(9), 1381-1401. doi:10.1080/09500693.2015.1036386
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Institute of Educational Research.
- Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, 45(3), 669-700. doi:10.3102/0002831208316955
- Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, 63(3), 66-70.
- Shouse, A. W., Schweingruber, H. A., & Duschl, R. A. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academy Press.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189. doi:10.3102/0034654307313795
- Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, 46(6), 610-631. doi:10.1002/tea.20313
- Steedle, J. T., & Shavelson, R. J. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching*, 46(6), 699-715. doi:10.1002/tea.20308
- Taylor, C. S., & Nolen, S. B. (2008). *Classroom assessment: Supporting teaching and learning in real classroom* (2nd ed.). Upper Saddle River, NJ: Pearson Education.

- Testa, I., Galano, S., Leccia, S., & Puddu, E. (2015). Development and validation of a learning progression for change of seasons, solar and lunar eclipses, and moon phases. *Physical Review Special Topics-Physics Education Research*, 11(2), 020102-1-020102-22. doi:10.1103/PhysRevSTPER.11.020102
- Treagust, D. F., Jacobwitz, R., Gallagher, J. L., & Parker, J. (2001). Using assessment as a guide in teaching for understanding: A case study of a middle school science class learning about sound. *Science Education*, 85(2), 137-157.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Soubberman, Eds. & Trans.). Cambridge, MA: Harvard University Press. (Original works published 1930-1935).
- Wang, J.-R., Kao, H.-L., & Lin, S.-W. (2010). Preservice teachers' initial conceptions about assessment of science learning: The coherence with their views of learning science. *Teaching and Teacher Education*, 26(3), 522-529. doi:10.1016/j.tate.2009.06.014
- Wang, T.-H., Wang, K.-H., & Huang, S.-C. (2008). Designing a web-based assessment environment for improving pre-service teacher assessment literacy. *Computers & Education*, 51(1), 448-462. doi:10.1016/j.compedu.2007.06.010
- Wang, T.-H., Wang, K.-H., Wang, W.-L., Huang, S.-C., & Chen, S.-Y. (2004). Web-based assessment and test analyses (WATA) system: Development and evaluation. *Journal of Computer Assisted Learning*, 20(1), 59-71. doi:10.1111/j.1365-2729.2004.00066.x
- Wang, W.-C. (2008). Assessment of differential item functioning. *Journal of Applied Measurement*, 9(4), 387-408.
- Warbruck, F., & Stachelscheid, K. (2010). Adaptive learning environment for chemical reactions. In G. Cakmakci & M. F. Tasar (Eds.), *Contemporary science education research: Learning and assessment* (pp. 425-429). Ankara, Turkey: Pegem Akademi.
- Wilson, M. (2008). Cognitive diagnosis using item response models. *Journal of Psychology*, 216(2), 74-88. doi:10.1027/0044-3409.216.2.74
- Wilson, M., & Carstensen, C. (2005). Assessment to improve learning in mathematics: The BEAR assessment system. *Journal of Educational Research and Development*, 1(3), 27-50.
- Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR assessment system. *Higher Education*, 52(4), 635-663. doi:10.1007/s10734-004-7263-y
- Wilson, M., Scalise, K., Reitze, A. M. A., Bein, E., Boussetot, T., Gochyyev, P., ...Vorp, R. (2010, May). *Progress monitoring for real classroom contexts: The formative assessment delivery system*. Paper presented at the annual meeting of the American Educational Research

Association, Denver, CO.

Wu, M. L., Adams, R. J., & Wilson, M. R. (2007). *ConQuest* [Computer software and manual]. Camberwell, Australia: Australian Council for Educational Research.

Journal of Research in Education Sciences

2017, 62(3), 1-23

doi:10.6209/JORIES.2017.62(3).01

Glimpse Into Teaching and Learning Using Rasch Analyses of a Teacher-Made Science Achievement Test

Ying-Tze Chen

Institute of Education,
National Sun Yat-sen University

Hsiao-Chi Ho

Centre for Humanities Innovation and
Social Practices, National Sun Yat-sen
University

Kun-Hsia Liu

Institute of Education,
National Sun Yat-sen University

Huann-Shyang Lin

Center for General Education,
National Sun Yat-sen University

Ying-Yao Cheng

Institute of Education,
National Sun Yat-sen University

Abstract

This study examined the quality of a teacher-made achievement test using Rasch analyses and provided useful assessment feedback information for improving teaching and testing. The participants were 8,936 seventh-grade students (4,617 boys and 4,319 girls) from the Kaohsiung area in Taiwan. Data were collected through a teacher-made biology achievement test and analyzed using the Rasch model. The results indicated that the test clearly discriminated students' abilities. Some questions, however, were found to have unfair content for girls. Additionally, some mismatches were identified between teachers' expectations and students' assessments of question difficulty. This paper also provides assessment feedback diagrams, including a person-item map and diagnostic map, and a method of identifying "the zone of fundamental learning development" and "the zone of proximal learning development" is proposed to enable teachers to plan students' learning progression.

Keywords: assessment feedback, learning progression, Rasch model, teacher-made achievement test, zone of proximal learning development

Corresponding Author: Ying-Yao Cheng, E-mail: chengyy@mail.nsysu.edu.tw

Manuscript received: Aug. 18, 2016; Revised: Oct. 13, 2016, Jan. 11, 2017; Accepted: Jan. 25, 2017.

